在线学习资料支持

您可以在华为企业业务网站获得E-Learning课程、培训教材 、产品资料、软件工具、技术案例等：

1、 E-Learning课程：登录*华为在线学习网站*，进入 “*华为培训/在线学习*” 栏目

免费E-Learning课：对网站所有用户免费开放

职业认证E-Learning课：通过任何一项职业认证即可学习所有职业认证培训E-Learning课程

渠道赋能E-Learning课：对华为企业业务合作伙伴免费开放

2、培训教材：登录*华为在线学习网站*，进入 “*华为培训/面授培训*” ， 在具体课程页面即可下载教材。

华为职业认证培训教材、华为产品技术培训教材。无需注册即可下载

3、华为在线公开课(LVC)： http://support.huawei.com/ecommunity/bbs/10154479.html

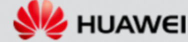
企业网络、UC&C、安全、存储等诸多领域的职业认证课程，华为讲师公开授课

4、产品资料下载： http://support.huawei.com/enterprise/#tabname=productsupport

5、软件工具下载：http://support.huawei.com/enterprise/#tabname=softwaredownload

更多内容请访问：

- http://learning.huawei.com/cn
- http://support.huawei.com/enterprise/
- http://support.huawei.com/ecommunity/



更多资料获取：http://learning.huawei.com/cn

华为数通认证系列教程-**HCDP IESN**

部署企业级交换网络

Implementing Enterprise Switching Networks



华为技术有限公司

更多资料获取：http://learning.huawei.com/cn

版权声明

版权所有 © 华为技术有限公司 2012。 保留一切权利。

本书所有内容受版权法保护，华为拥有所有版权，但注明引用其他方的内容除外。未经华为技术有限公司事先书面许可，任何人、任何组织不得将本书的任何内容以任何方式进行复制、经销、翻印、存储于信息检索系统或使用于任何其他任何商业目的。

版权所有 侵权必究。

商标声明

HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

华为数通认证系列教程-HCDP-Enterprise

华为认证数据通信资深工程师-企业级

1.6 版本

更多资料获取：http://learning.huawei.com/cn

华为认证体系介绍

依托华为公司雄厚的技术实力和专业的培训体系，华为认证考虑到不同客户对ICT技术不同层次的需求，致力于为客户提供实战性、专业化的技术认证。

根据ICT技术的特点和客户不同层次的需求，华为认证为客户提供面向十三个方向的四级认证体系。

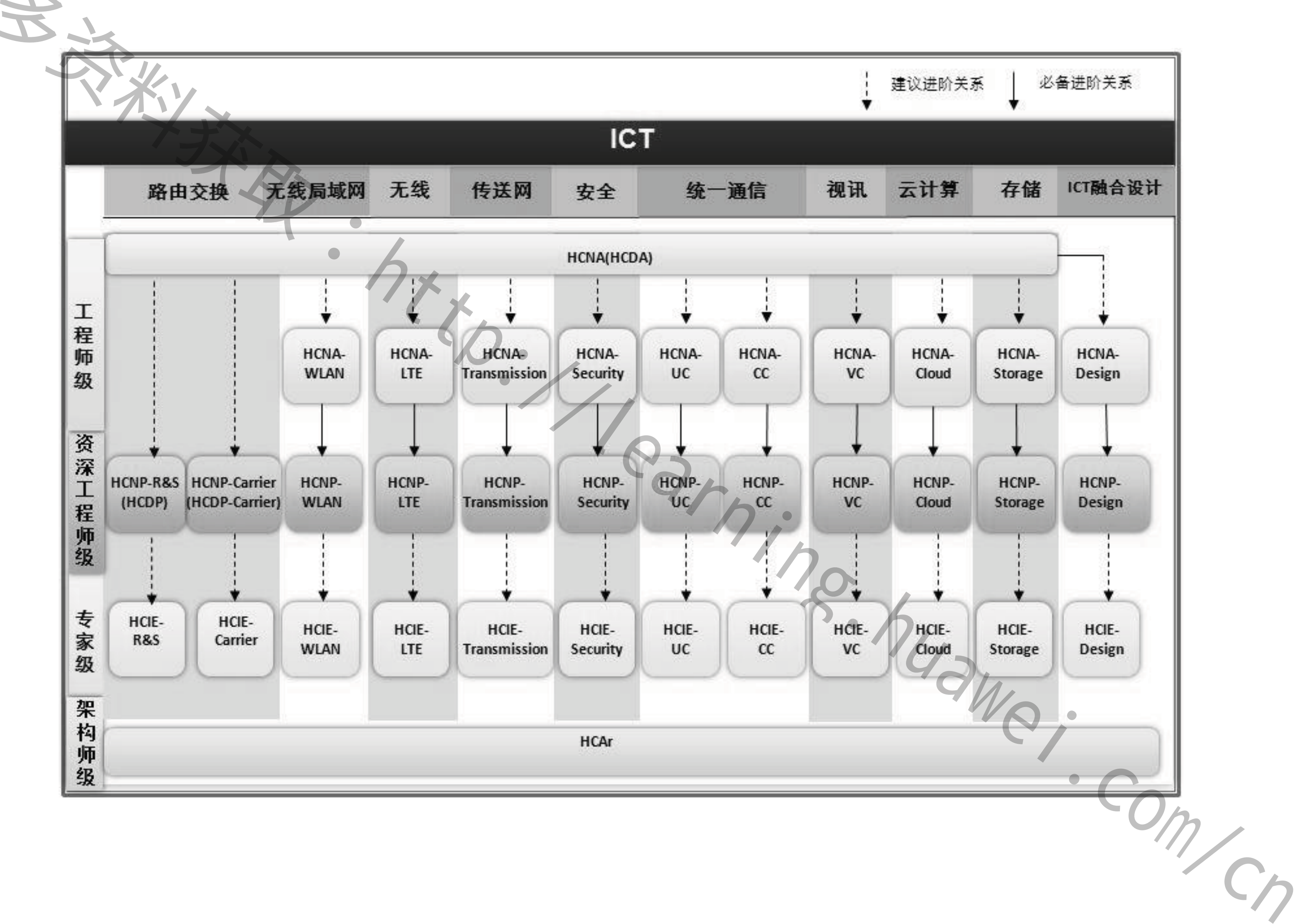
HCNA(HCDA)认证定位于中小型网络的基本配置和维护。HCNA(HCDA)认证包括但不限于：网络基础知识；流行网络的基本连接方法；基本的网络建造；基本的网络故障排除；华为路由交换设备的安装和调试。通过HCNA(HCDA)认证，将证明您对中小型网络有初步的了解，了解面向中小型企业的网络通用技术，并具备协助设计中小企业网络以及使用华为路由交换设备实施设计的能力。拥有通过HCNA(HCDA)认证的工程师，意味着中小企业有能力完成基本网络搭建，并将基本的语音、无线、云、安全和存储集成到网络之中，满足各种应用对网络的使用需求。

HCNP-Enterprise (HCDP-Enterprise)认证定位于中小型网络的构建和管理。HCNP-Enterprise (HCDP-Enterprise)认证包括但不限于：网络基础知识；交换机和路由器原理；TCP/IP协议簇；路由协议；访问控制；网络故障的排除；华为路由交换设备的安装和调试。通过HCNP-Enterprise (HCDP-Enterprise)认证，将证明您对中小型网络有全面深入的了解，掌握面向中小型企业的网络通用技术，并具备独立设计中小企业网络以及使用华为路由交换设备实施设计的能力。拥有通过HCNP-Enterprise (HCDP-Enterprise)认证的工程师，意味着中小企业有能力完成完整网络的搭建，将企业中所需的语音、无线、云、安全和存储全面地集成到网络之中，并且能满足各种应用对网络的使用需求，进而提供较高的安全性、可用性和可靠性。

HCIE-Enterprise 认证定位于大中型复杂网络的构建、优化和管理。HCIE-Enterprise认证包括但不限于：不同网络和各种路由器交换机之间的互联；复杂连接问题的解决；使用技术解决方案提高带宽、缩短相应时间、最大限度地提高性能、加强安全性和支持全球应用；复杂网络的故障排除。通过HCIE-Enterprise认证，将证明您对大型网络有全面深入的了解，掌握面向大型企业网络的技术，并具备独立设计各种企业网络以及使用华为路由交换设备实施设计的能力。拥有通过HCIE-Enterprise认证的工程师，意味着大中企业有能力独立完成完整的网络搭建，将企业中所需的语音、无线、云、安全和存储全面地集成到网络之中，并且能满足各种应用对网络的使用需求；能够提供完整的故障排除能力；能根据企业和网络技术的发展，规划企业网络的发展，并提供高安全性、可用性和可靠性。

华为认证协助您打开行业之窗，开启改变之门，屹立在ICT世界的潮头浪尖！

更多资料访问：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

前言

简介

本书为 HCDP-IESN 认证培训教程，适用于准备参加 HCDP-IESN 考试的学员或者希望系统掌握通用交换网协议原理以及在华为通用路由平台 VRP 上的实现的读者。

内容描述

本书共包含四个 Module，由浅入深地介绍了通用交换技术协议原理、MPLS 技术协议原理以及在华为 VRP 上的配置与实现；同时重点介绍华为以太网交换机产品及网络运用。

Module 1 详细介绍了 VLAN、GVRP 原理及其实现，包括 VLAN 二层互通和三层路由以及 VLAN 嵌套（QinQ）等，帮助读者全面深入地了解 VLAN 工作原理及其在 VRP 上的配置；

Module 2 对 STP 协议进行了详细的描述，包括 STP、RSTP、MSTP 三部分内容，使读者掌握 STP 协议工作原理以及在 VRP 上的实现；

Module 3 详细介绍了各种接入技术原理和配置，包括 802.1x、DHCP 和 RRPP，帮助读者对接入技术协议有一个全面的了解；

Module 4 主要介绍 MPLS、LDP 协议基本原理等，使读者熟悉 MPLS。

本书引导读者全面掌握企业级交换技术及其在华为产品中的实现，读者也可以根据自身情况选择感兴趣的章节阅读。

读者必备知识背景

为了更好地掌握本书内容，阅读本书的读者应首先具备以下基本条件之一：

(1) 参加过 HCDA 培训

(2) 通过 HCDA 考试

(3) 熟悉 TCP/IP 协议栈和以太网基础知识

(4) 熟悉以太网交换机基本工作原理

更多资料获取：http://learning.huawei.com/cn

本书常用图标



IPv6路由器



SOHO路由器



语音模块的路由器


中低端路由器

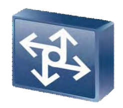

高端路由器


核心路由器


集线器


插座式交换机


汇聚交换机


核心交换机


边缘交换机


堆叠交换机


AP


AP大功率


无线网桥


无线网卡


接入服务器


语音网关


防火墙


网络电话系统

更多资料获取：http://learning.huawei.com/cn

目 录

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn

Module 1

VLAN

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

虚拟局域网（VLAN）技术为以太网引入了灵活的控制手段，被广泛应用。

本文旨在介绍VLAN及相关技术的基本原理和实现。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 了解VLAN的基本原理及基本配置
- 了解VLAN相关技术的基本原理及配置

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page2

HUAWEI

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

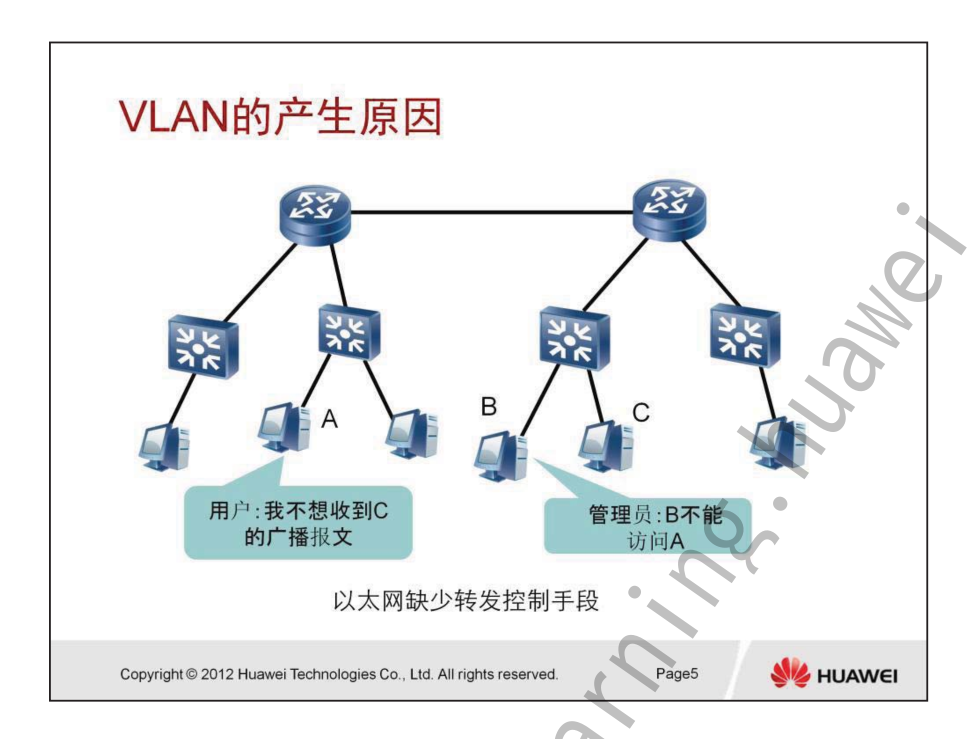


目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page4 HUAWEI



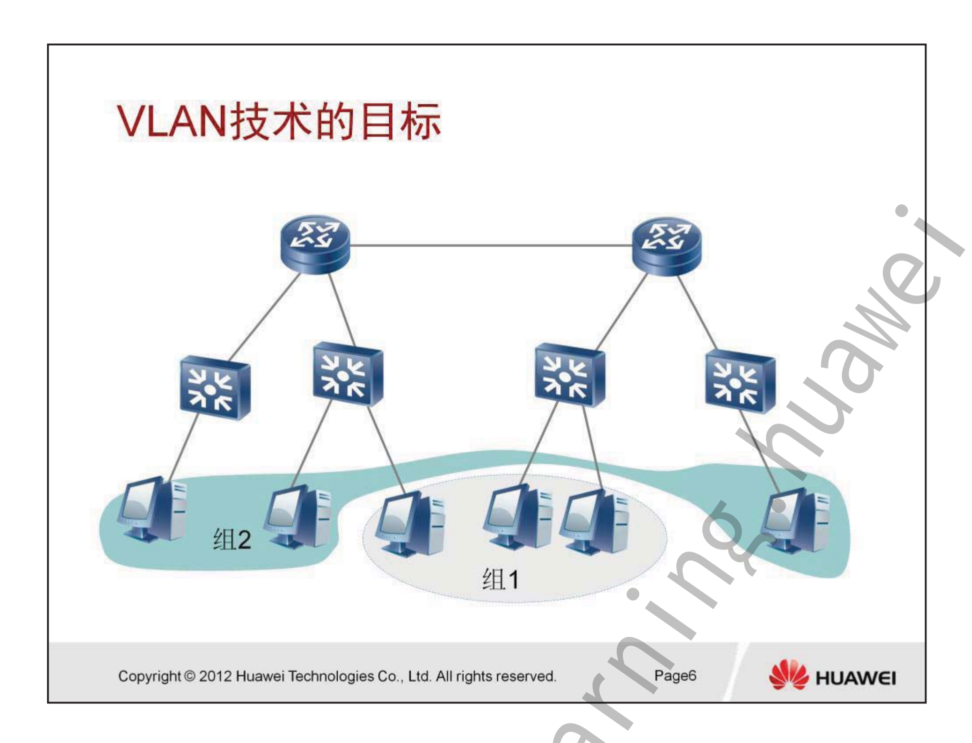
更多资料获取：http://learning.huawei.com/cn



传统的以太网交换机在转发数据时，采用源地址学习的方式，自动学习各个端口连接的主机的MAC地址，形成MAC地址表，然后依据此表进行以太网帧的转发。整个转发的过程自动完成，所有端口都可以互访，维护人员无法控制端口之间的转发，例如B主机不能访问A主机就无法实现。该网络存在如下缺陷：

- 网络的安全性差。由于各个端口之间可以直接互访，降低了网络的安全性。
- 网络效率低。用户可能收到大量不需要的报文，例如不必要的广播报文，这些报文同时消耗网络带宽资源和客户主机CPU资源。
- 业务扩展能力差。网络设备平等的对待每台主机的报文，无法实现有差别的服务，例如无法优先转发用于网络管理的以太网帧。

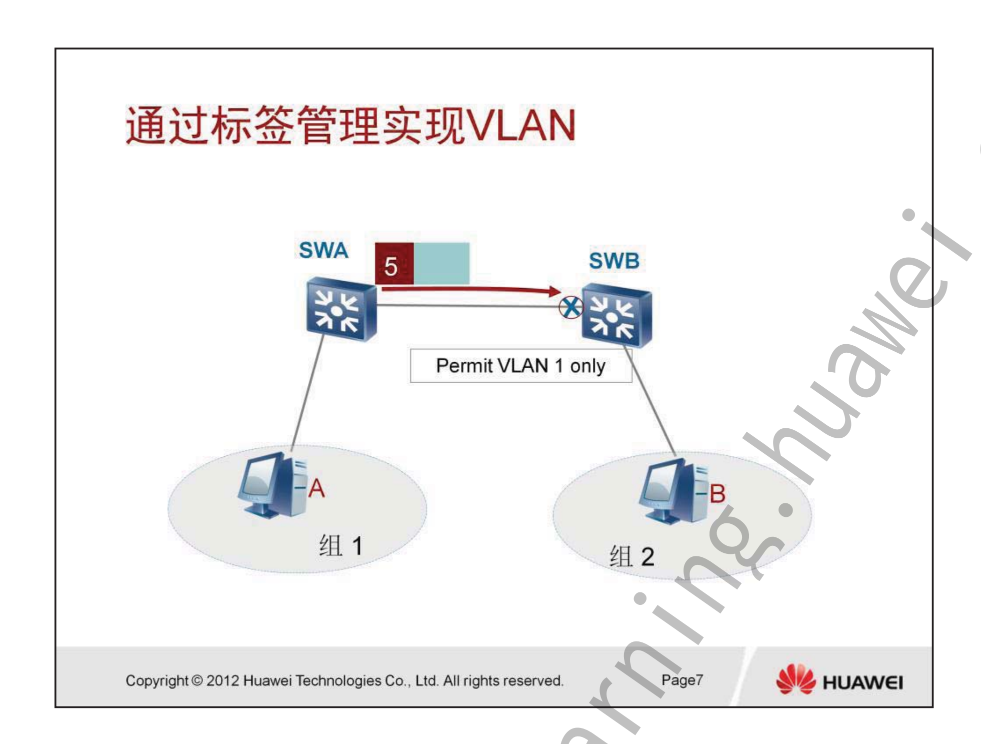
更多资料获取：http://learning.huawei.com/cn



VLAN技术把用户划分成多个逻辑的网络（group），组内可以通信，组间不允许通信，二层转发的单播、组播、广播报文只能在组内转发。同时，VLAN技术可以很容易地实现组成员的添加或删除。

即VLAN技术提供了一种管理手段，控制终端之间的互通。如上图，组1和组2的PC无法相互通信。

更多资料获取：http://learning.huawei.com/cn

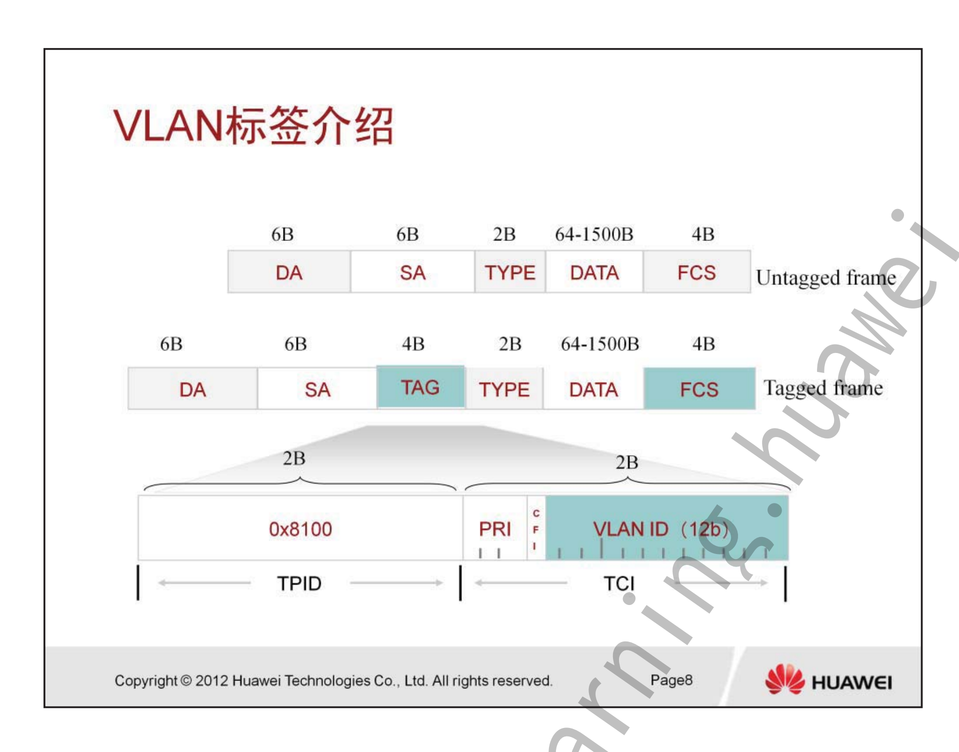


为了实现转发控制，在待转发的以太网帧中添加VLAN标签，然后设定交换机端口对该帧和标签的处理方式，包括丢弃帧、转发帧、添加标签、移除标签。

转发帧时，检查以太网报文中携带的VLAN标签，是否为该端口允许通过的标签，判断出该以太网帧是否能够从端口转发出去。上图中，假设有一种方法，SWA将主机A发出的所有以太网帧都加上标签5，此后查询二层转发表，根据目的MAC地址将该帧转发到SWB连接的端口。由于在该端口配置了仅允许VLAN 1通过，主机A发的帧将被丢弃。

即支持VLAN技术的交换机，转发以太网帧时不再仅仅依据目的MAC地址，同时还要考虑该端口的VLAN配置情况，从而实现对二层转发的控制。

更多资料获取：http://learning.huawei.com/cn



DA: Destination address;

SA: Source Address

TAG:VLAN TAG

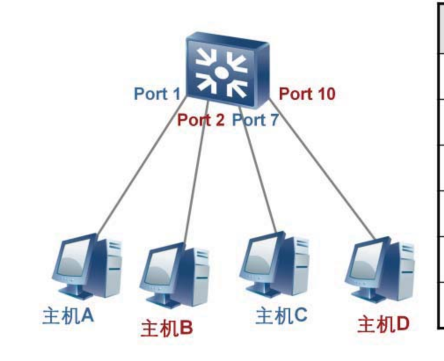
TPID：Tag Protocol Identifier，2字节，固定取值，0x8100，是IEEE定义的新类型，表明这是一个携带802.1Q标签的帧。

TCI：Tag Control Information，2字节。帧的控制信息，详细说明如下：

- Priority：3比特，表示以太网帧的优先级。一共有8种优先级，0－7，用于提供有差别的转发服务。
- CFI：Canonical Format Indicator，1比特。用于令牌环/源路由FDDI介质访问中指示地址信息的比特次序信息，即先传送的是低比特位还是高比特位。
- VLAN Identified：VLAN ID，12比特，取值从0到4095。

更多资料获取：http://learning.huawei.com/cn

如何生成VLAN标签



端口	PVID
Port1	5
Port2	10
……	……
Port7	5
……	……
Port10	10

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.
Page9


所有以太网帧在交换机中都是以tagged frame的形式流动的，即某端口从对端设备收到的帧，有可能是untagged的，但是从本交换机其它端口转发来的帧，一定是tagged的。如果收到的是tagged frame，则进入转发过程，如果该端口收到的是untagged frame，则必须加上标签。以下几种方法可以确定标签中的VLAN ID取值：

- 基于端口：网络管理员给交换机的每个端口配置PVID，即Port VLAN ID，有些场合称为端口默认VLAN。如果收到的是untagged帧，则VLAN ID的取值为PVID。
- 基于MAC地址：网络管理员配置好MAC地址和VLAN ID的映射关系表，如果收到的是untagged帧，则依据该表添加VLAN ID。
- 基于协议：网络管理员配置好以太网帧中的协议域和VLAN ID的映射关系表，如果收到的是untagged帧，则依据该表添加VLAN ID。
- 基于子网：根据报文中的IP地址信息，确定添加的VLAN ID。
- 基于策略：根据上面几种划分依据组合进行划分。

如果设备同时支持多种方式，一般情况下，优先使用顺序为：基于策略－基于子网－基于协议－基于MAC地址－基于端口。目前常用的是基于端口的方式。

更多资料获取：http://learning.huawei.com/cn

VLAN划分方式

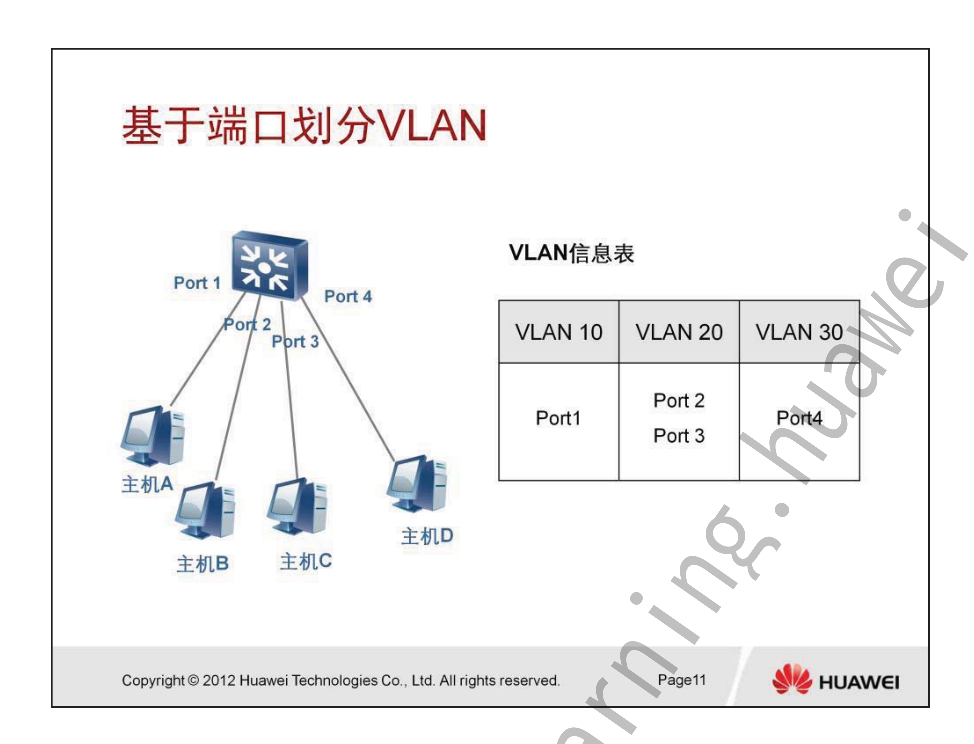
- 根据端口划分: 基于端口的VLAN
- 根据MAC划分: 基于MAC的VLAN
- 根据IP进行划分: 基于IP子网的VLAN
- 根据协议划分: 基于协议的VLAN
- 根据几种划分依据组合进行划分: 基于策略的VLAN

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page10



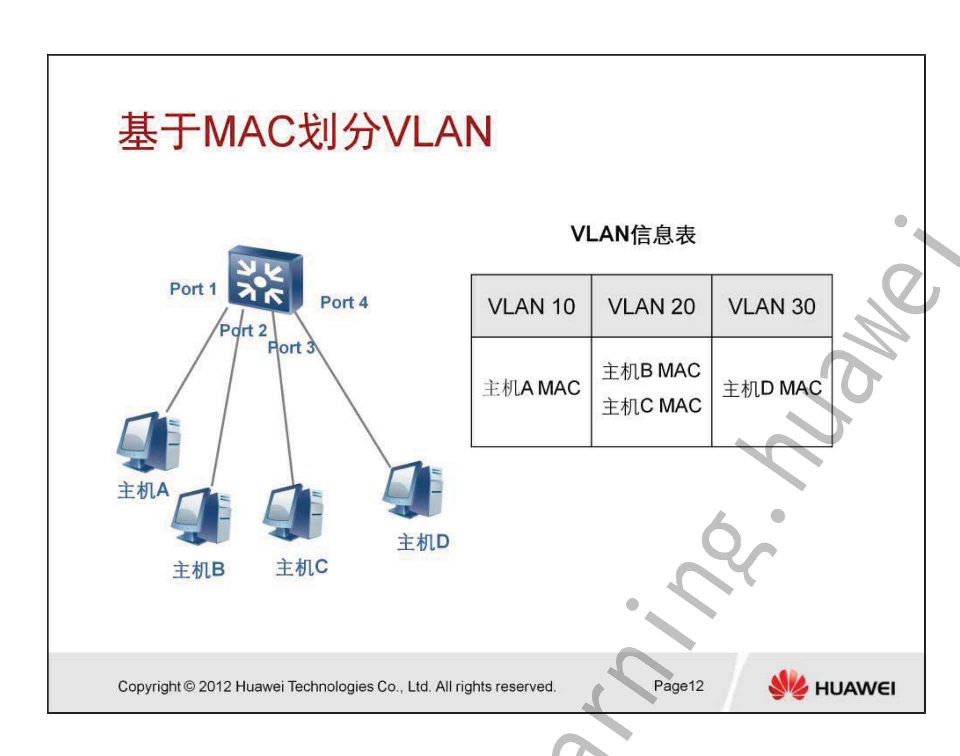
更多资料获取：http://learning.huawei.com/cn





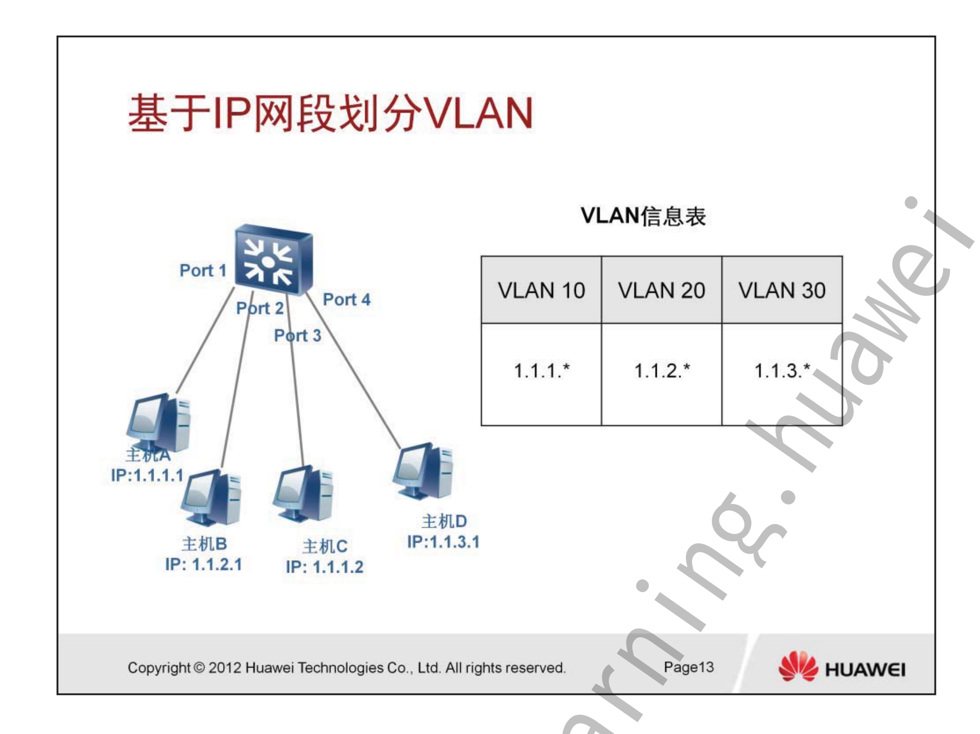
基于端口划分VLAN，是最简洁、最广泛使用的划分方式，可以把多个端口划入一个VLAN 。

更多资料获取：http://learning.huawei.com/cn



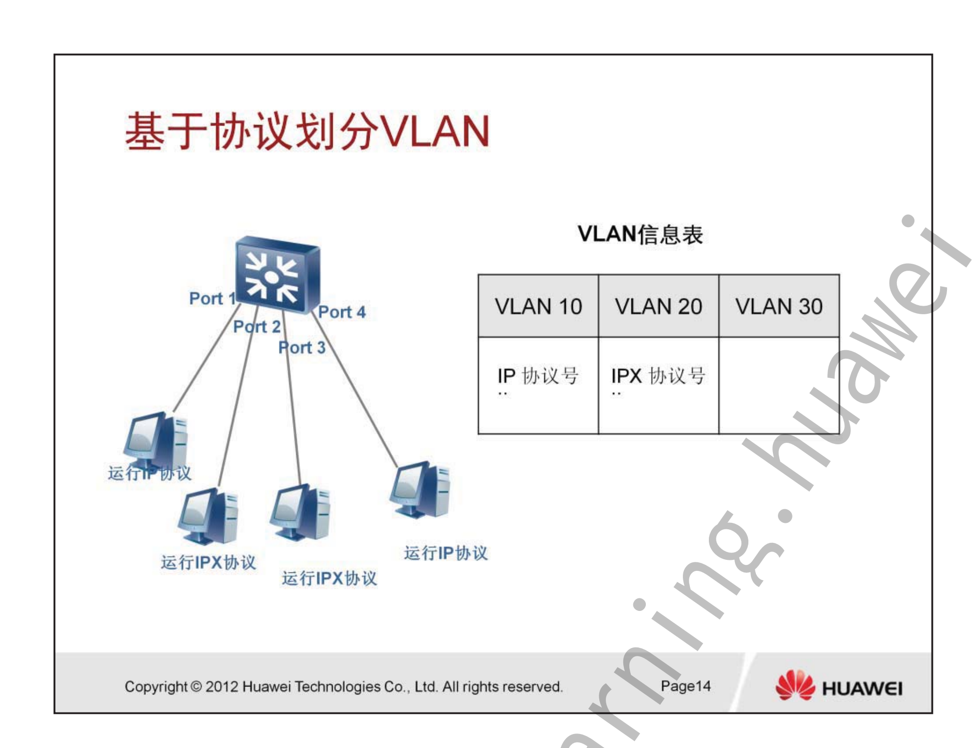
基于大的MAC地址划分VLAN：交换机根据报文的源MAC来确定报文应该在哪个VLAN中进行转发。实际上就是根据终端设备的MAC来划分VLAN。

更多资料获取：http://learning.huawei.com/cn



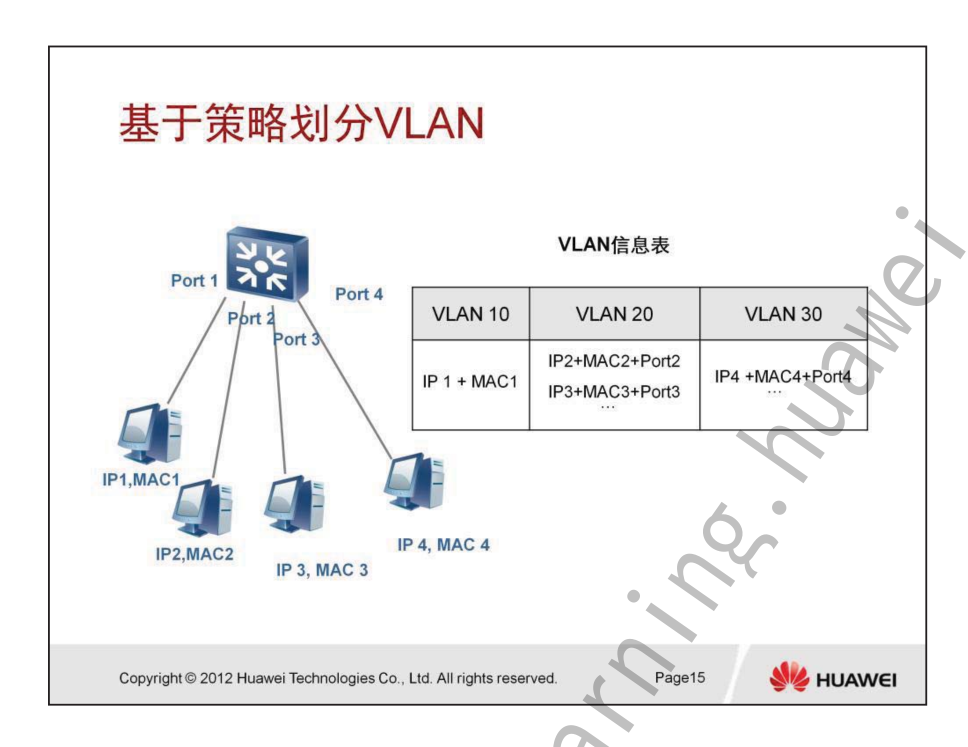
- 基于IP网段划分VLAN，即根据报文源IP及掩码来确定报文所属VLAN。比如，可以配置1.1.1.0/24 加入VLAN 10；1.1.2.0/24加入VLAN 20；1.1.3.0/24加入VLAN 30；再配置某个端口属于这些VLAN。在这个端口上，对于收到的不带VLAN标签的报文：源ip在1.1.1.0/24内的报文将被打上VLAN10标签。在1.1.2.0/24内的将被打上VLAN20标签。在1.1.3.0/24内的将被打上VLAN30标签。

更多资料获取：http://learning.huawei.com/cn



协议VLAN即根据端口接收到的报文所属的协议（族）类型及封装格式来给报文分配不同的VLAN ID。如IP、IPX、AppleTalk协议族；Ethernet II，802.3，802.3/802.2 LLC， 802.3/802.2 SNAP等封装格式。设备对端口上收到的untagged的报文，判断其协议类型，打上对应的VLAN标签，并在这个VLAN内转发。

更多资料获取：http://learning.huawei.com/cn



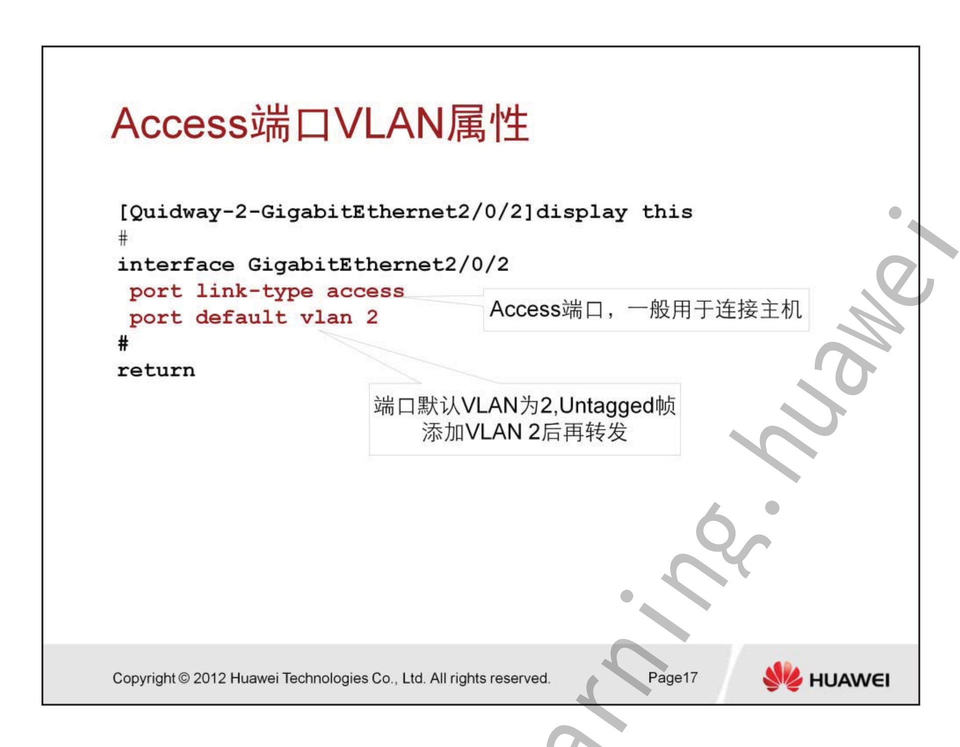
策略VLAN:对于匹配了策略IP+MAC或者IP+MAC+PORT的Untagged报文进行VLAN划分。

更多资料获取：http://learning.huawei.com/cn



引入VLAN功能后，交换机的端口被划分为3种类型：Acess端口、Trunk端口、Hybrid端口。

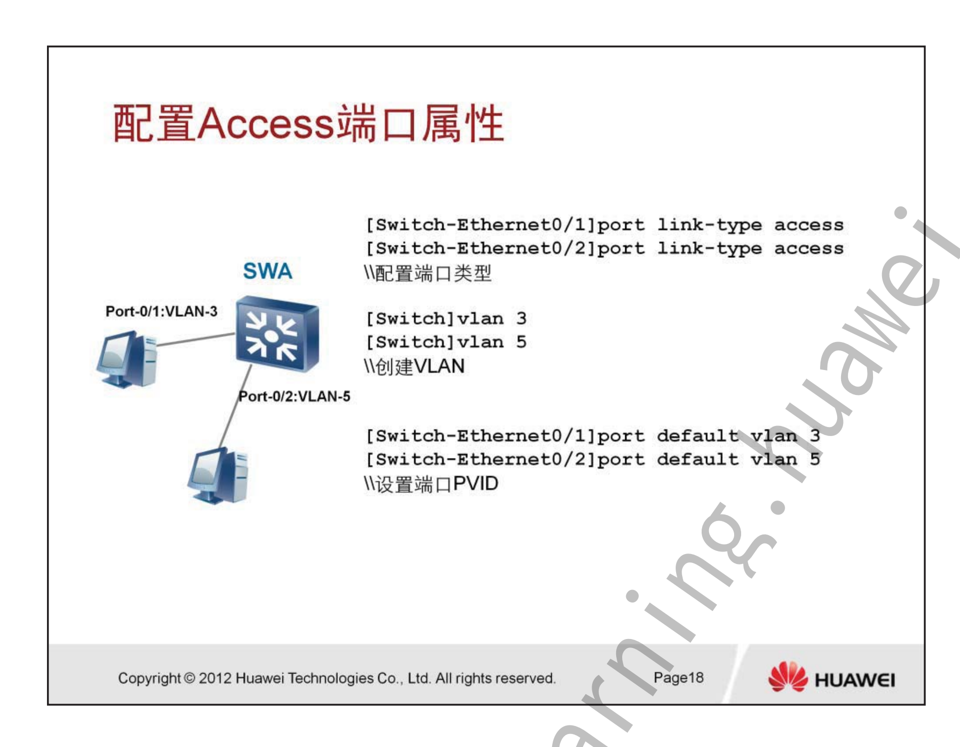
更多资料获取：http://learning.huawei.com/cn



Access端口，用于连接主机，有如下特点：

- 仅仅允许VLAN ID与端口的PVID相同的数据帧通过本端口
- 如果该端口收到的对端设备发送的帧是untagged，交换机将强制加上该端口的PVID
- Access端口发往对端设备的以太网帧永远是untagged frame
- 很多型号的交换机默认端口类型是access，PVID默认是1，VLAN 1由系统创建，不能被删除。

更多资料获取：http://learning.huawei.com/cn



可以通过下述方法设置access端口PVID：

在创建VLAN后，将access端口加入到该VLAN中

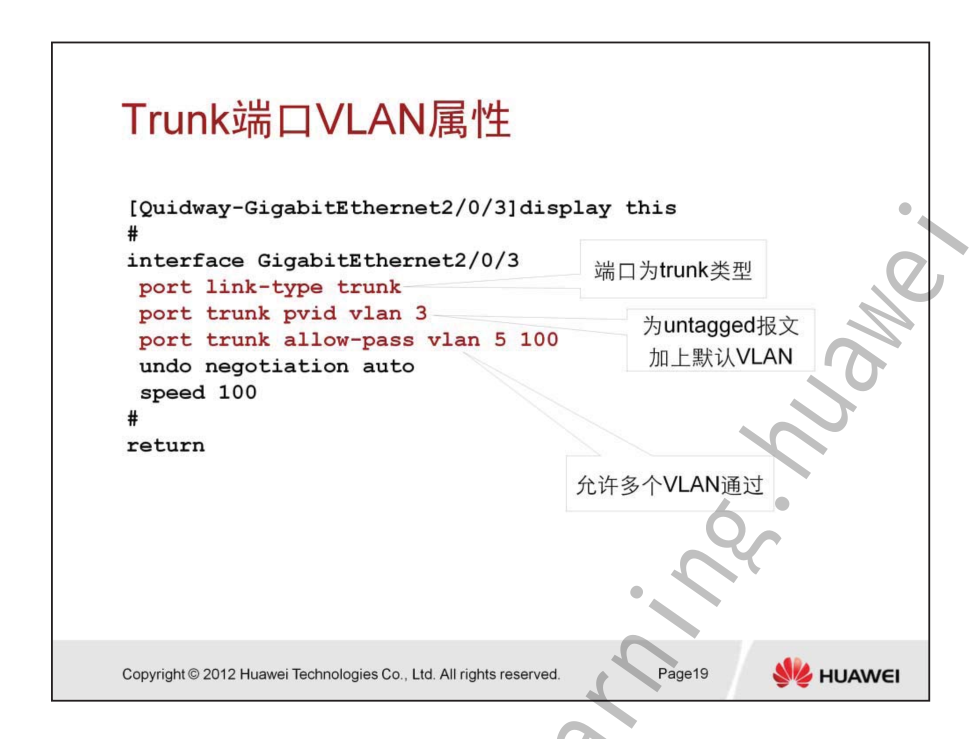
[Switch]vlan 3

[Switch-vlan3]port ethernet 0/1

[Switch]vlan 5

[Switch-vlan5]port ethernet 0/2

更多资料获取：http://learning.huawei.com/cn



Trunk端口：用于连接交换机，在交换机之间传递tagged frame，可以自由设定允许通过的多个VLAN ID，这些ID可以与PVID相同，也可以不同。

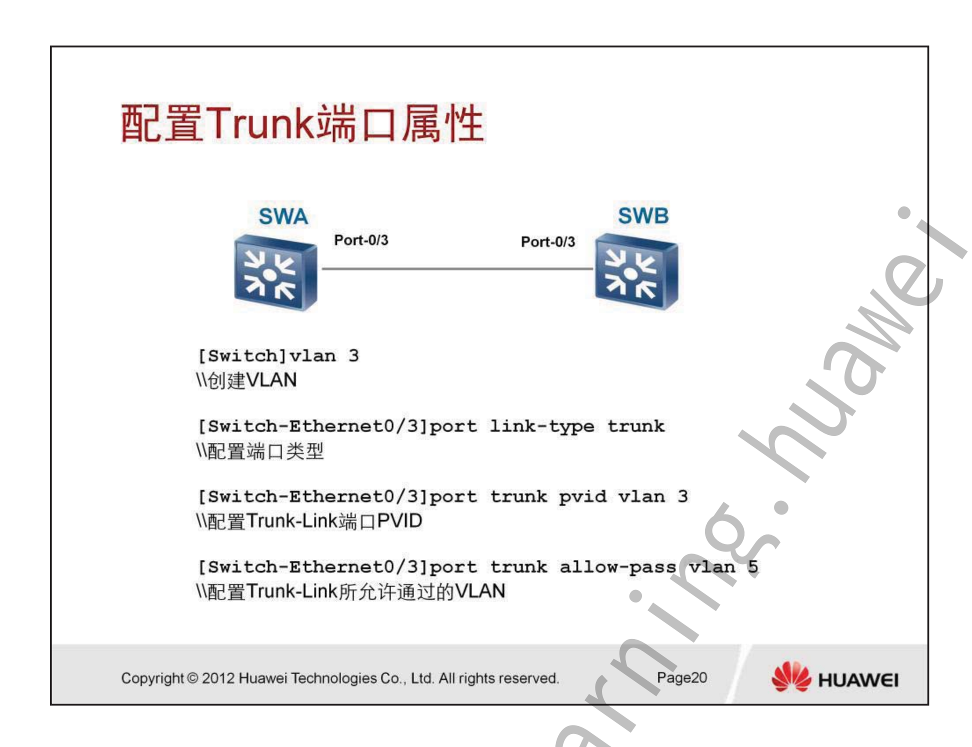
Trunk端口发送Tagged frame向其它设备时，规则如下：

该Tagged frame中的VLAN ID取值未出现在VLAN permitted列表中，则丢弃该帧。如果在列表中，则：

- 该Tagged frame中的VLAN ID取值与本端口的PVID相同，则移除标签后发往其它设备。由于每个端口的PVID取值是唯一的，因此仅仅在这一种情况下，Trunk端口发往其它设备的帧是untagged frame；
- 该Tagged frame中的VLAN ID取值与本端口的PVID不同，则不做任何改变，发往对端设备。

VLAN passing：一般情况下，VLAN passing与VLAN permited查询内容相同。但是通过GVRP协议动态注册的VLAN，如果未在端口进行注册，则该VLAN ID不会出现在VLAN passing列表中，相应的帧也不能从该端口转发出去。

更多资料获取：http://learning.huawei.com/cn



如图所示，可以通过以下命令配置Trunk端口属性：

\\创建VLAN

[Switch]vlan 3

\\配置端口类型

[Switch-Ethernet0/3]port link-type trunk

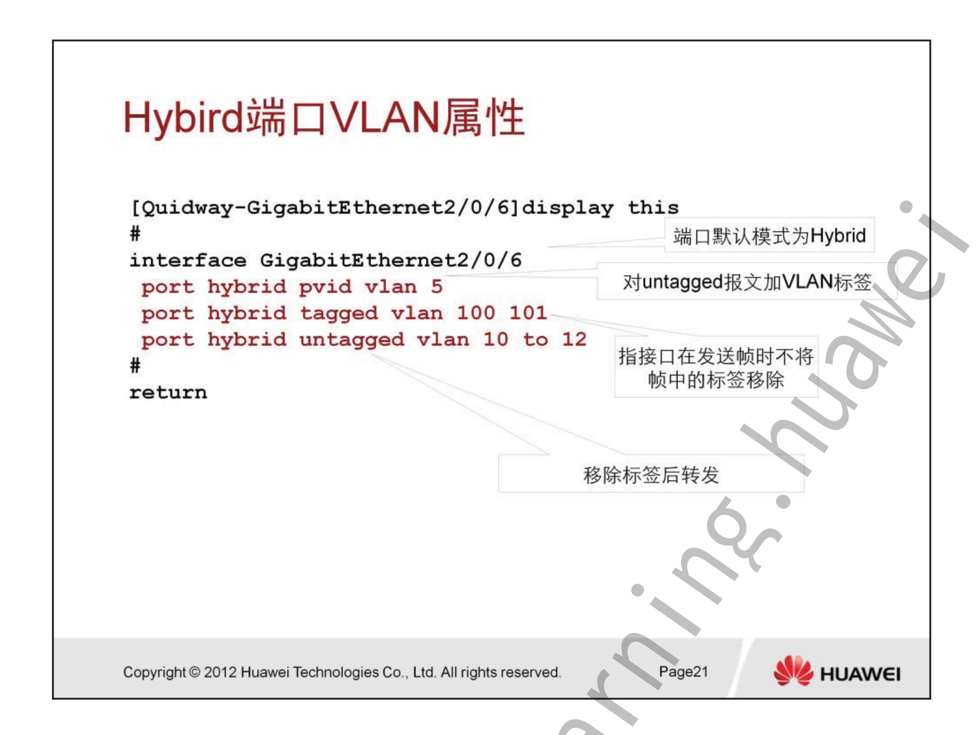
\\配置Trunk-Link端口PVID

[Switch-Ethernet0/3]port trunk pvid vlan 3

\\配置Trunk-Link所允许通过的VLAN（permitted VLAN）

[Switch-Ethernet0/3]port trunk permit vlan 5

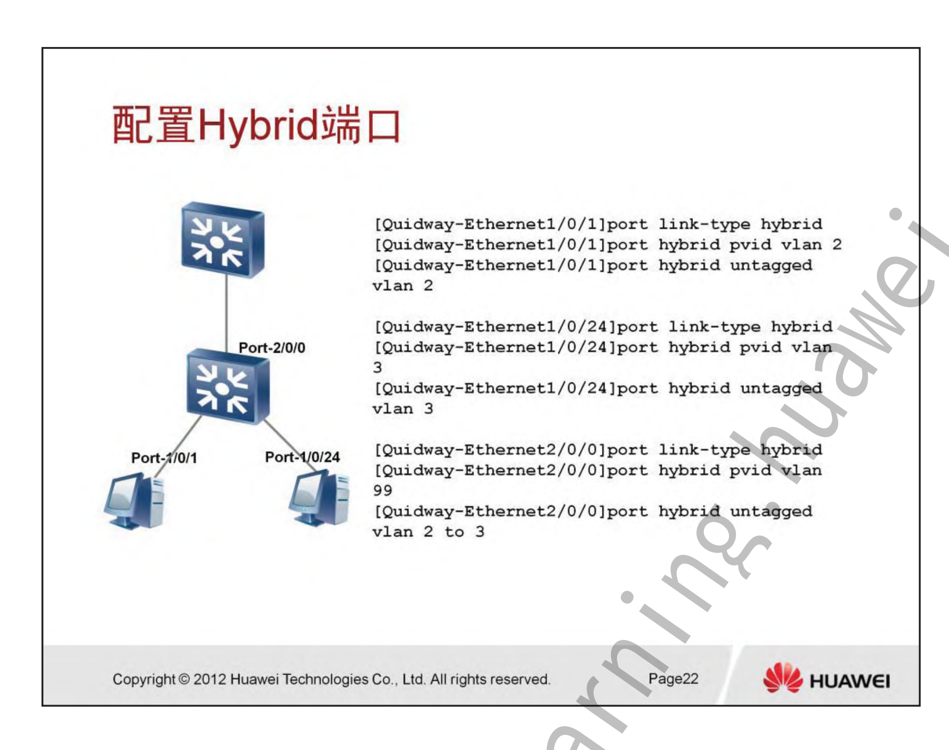
更多资料获取：http://learning.huawei.com/cn



Access端口发往其他设备的报文，都是untagged frame，而trunk端口仅在一种特定情况下才能发出untagged frame，其它情况发出的都是tagged frame。某些应用中，可能希望能够灵活的控制VLAN标签的移除。例如，在本交换机的上行设备不支持VLAN的情况下，希望实现各个用户端口相互隔离。

Hybrid端口可以灵活的控制VLAN标签的移除情况。例如如果待转发的帧中的VLAN ID是3，则按照trunk端口的转发模式转发；如果是4，则移除标签4后再转发。

更多资料获取：http://learning.huawei.com/cn



如果某Hybrid端口的tagged VLAN是none，而untagged VLAN只有一个取值，则该端口与access端口功能相同。同理，如果端口没有配置untagged VLAN，则与trunk端口功能相同。

如果将交换机的所有接入口设置不同的VLAN，例如2，3……24，上行口Untagged VLAN ID列表为2，3……24，则可以实现用户侧相互隔离，提高安全性，而上行帧为untagged frame，满足与上行设备互通的要求。

上图中的配置可以实现端口1/0/1和端口1/0/24的隔离，但是都可以与上行设备通信，且上行帧是untagged frame。

\\ tagged VLAN 配置示例

[Quidway-Ethernet2/0/0] port hybrid tagged vlan 3

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

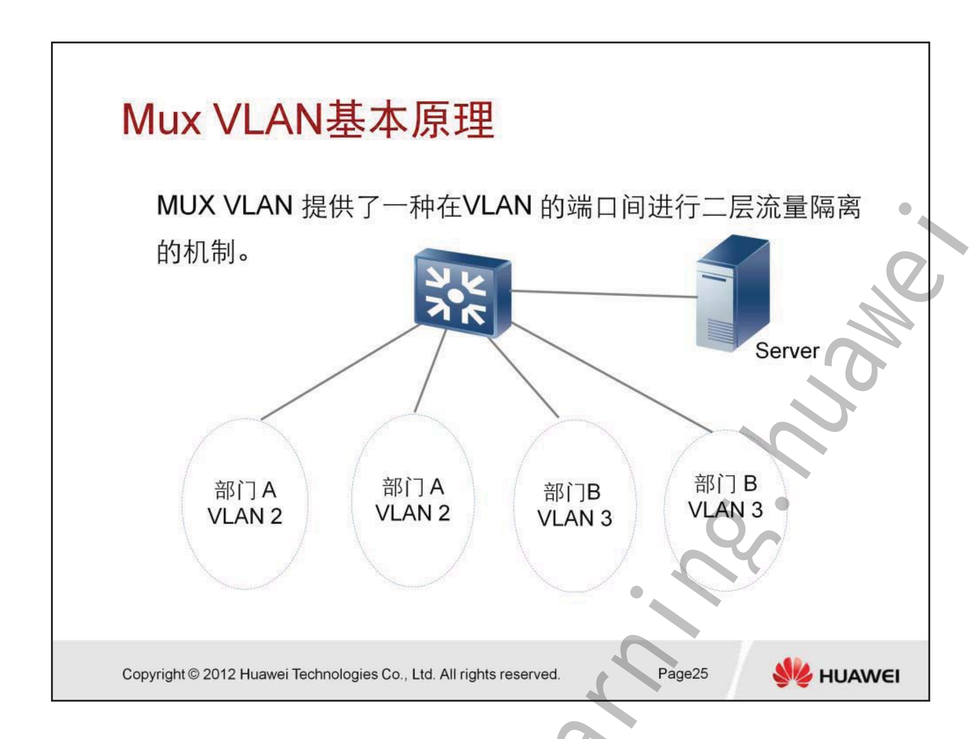


目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page24 HUAWEI



更多资料获取：http://learning.huawei.com/cn



比如在企业网络中，客户端口可以和服务器端口通讯，但客户端口之间不能互相通讯。

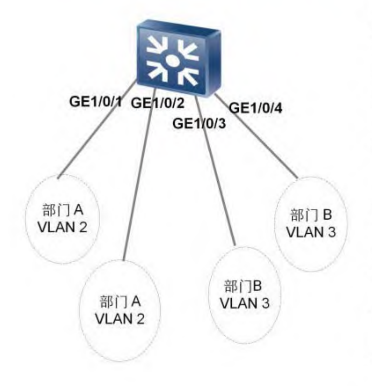
MUX VLAN 分为主VLAN 和从VLAN，从VLAN 又分为互通型从VLAN 和隔离型从VLAN。

主VLAN与从VLAN之间可以相互通信；互通型从VLAN内的端口之间可以互相通信，隔离型从VLAN内的端口之间不能互相通信，不同从VLAN之间不能互相通信。

部门A的员工之间不能互访，而部门B的员工之间可以互访，但是部门A和部门B不能互访，而公司所有员工均可以访问公司的服务器。对于该应用，可以使用MUX VLAN 来实现。可以将部门A的员工加入到隔离型从VLAN，而将部门B的员工加入互通型从VLAN，服务器加入主VLAN
。

更多资料获取：http://learning.huawei.com/cn

Mux VLAN 配置

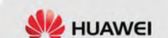



```
[Quidway]vlan batch 2 3 10
//创建VLAN
[Quidway]vlan 10
[Quidway-vlan10]mux-vlan
//配置主VLAN
[Quidway-vlan10]subordinate group 3
//配置MUX VLAN中的互通型从VLAN

[Quidway-vlan10]subordinate separate 2
//配置MUX VLAN中的隔离型从VLAN
[Quidway]interface gigabitethernet1/0/1
[Quidway-GigabitEthernet1/0/1]port mux-vlan enable
[Quidway-GigabitEthernet1/0/2]port mux-vlan enable
[Quidway-GigabitEthernet1/0/3]port mux-vlan enable
[Quidway-GigabitEthernet1/0/4]port mux-vlan enable
//使能端口下的MUX-VLAN功能
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page26



更多资料获取：http://learning.huawei.com/cn



ARP Proxy概述

ARP（Address Resolution Protocol）用于将一个IP地址映射到正确的MAC地址。

一个物理网络的子网（Subnet）中的源主机向另一个物理网络的子网中的目的主机发ARP request，和源主机直连的网关用自己接口的MAC地址代替目的主机回ARP reply，这个过程称为ARP 代理。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page27



ARP Proxy的基本过程如下：

源主机向另一物理网络的子网的目的主机发ARP请求；

与源主机网络相连的网关已经使能ARP PROXY功能，如果存在到达目的主机的正常路由，则代替目的主机以自己接口的MAC地址回应；

源主机向目的主机发送的IP报文都发给了路由器；

路由器对报文做正常的IP路由转发；

发往目的主机的IP报文通过网络，最终到达目的主机。

更多资料获取：http://learning.huawei.com/cn

ARP Proxy基本原理

当主机上没有配置缺省网关地址,它可以发送一个ARP请求,请求目的主机的MAC地址。使能ARP Proxy功能的交换机收到这样的请求后，会使用自己的MAC地址作为该ARP请求的回应，使得处于不同物理网络但网络号相同的主机之间可以正常的相互通信。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn

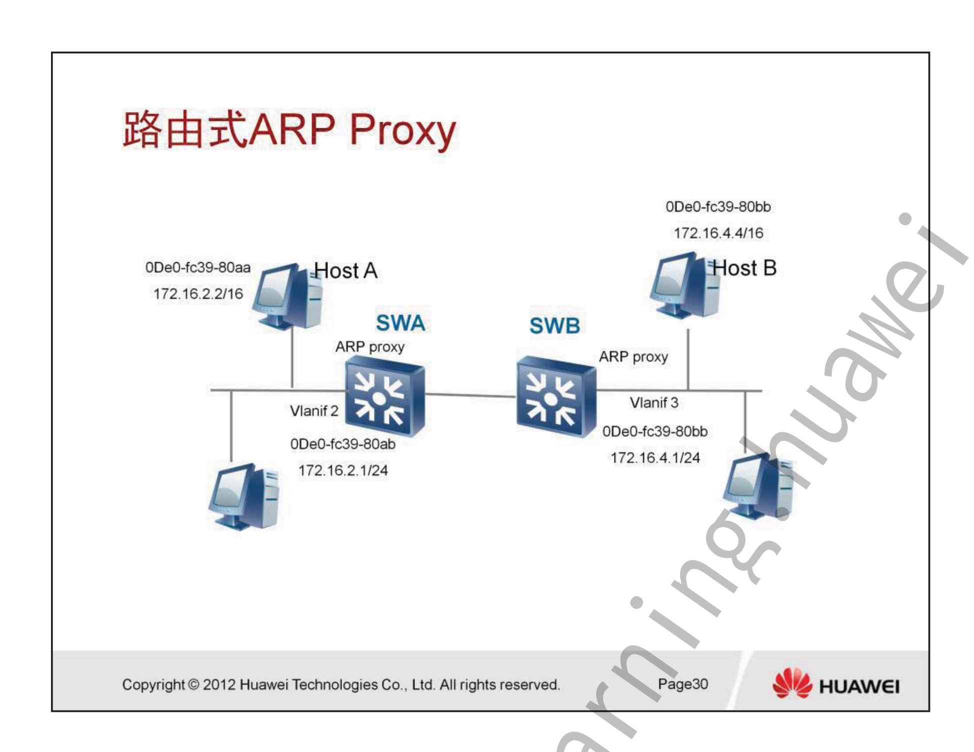
ARP Proxy方式

ARP Proxy方式	解决的问题
路由式ARP Proxy	解决同一网段不同物理网络上计算机的互通问题。
VLAN 内ARP Proxy	解决相同VLAN 内，且VLAN 配置用户隔离后的网络上计算机互通问题。
VLAN 间ARP Proxy	解决不同VLAN 之间对应计算机的三层互通问题。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page29



更多资料获取：http://learning.huawei.com/cn



路由式 ARP Proxy 就是使那些在同一网段却不在同一物理网络上的计算机或交换机能够相互通信的一种功能。

在实际的应用中，如果连接交换机的主机上没有配置缺省网关地址，数据将无法转发。

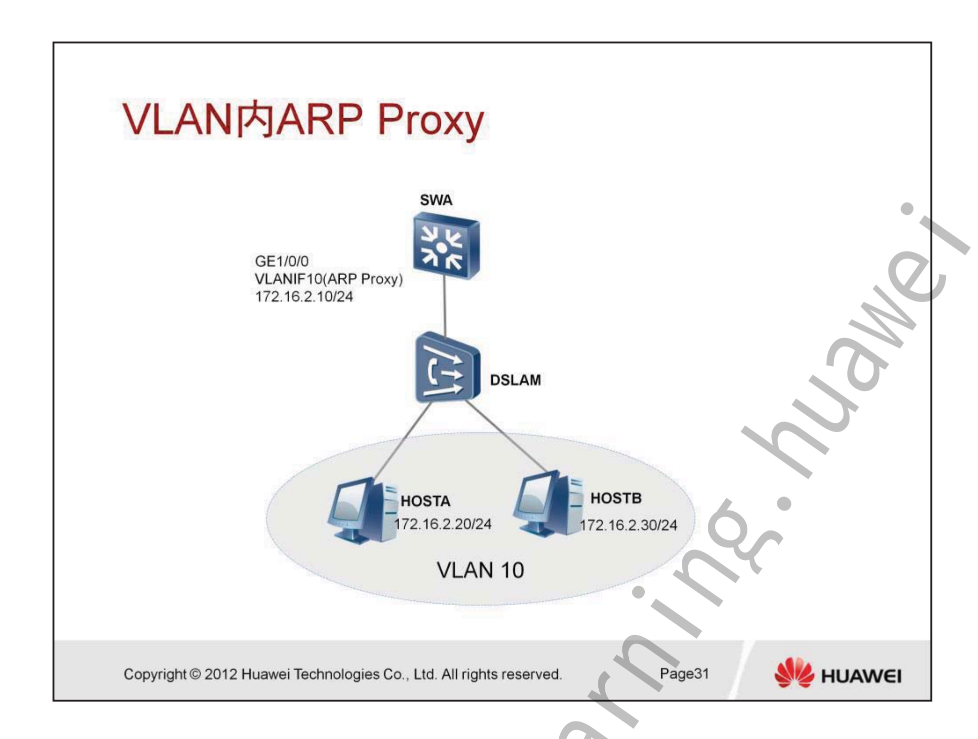
路由式ARP Proxy可以解决这个问题，主机发送一个ARP 请求（请求目的主机的MAC地址)，

使能ARP Proxy功能的交换机收到这样的请求后，会使用自己的MAC地址作为该ARP请求的回应，以此来欺骗主机进行数据转发。

使能ARP Proxy功能的交换机还可隐藏物理网络的细节，使得处于不同物理网络但网络号相同的Ethernet A 和Ethernet B 的内部主机之间可以正常的相互通信。

由于HostA 只有16 位掩码，它认为自己直连到172.16.0.0 网段，当HostA 需要与172.16.0.0 网段上的设备例如HostB 通信时，它使用ARP 请求HostB 的物理地址。由于交换机不转发广播，HostB 收不到HostA 发出的ARP 请求。如果在SwitchA 的接口VLANIF2 上使能路由式Proxy ARP，由SwitchA 转发HostA 与HostB 之间的IP 报文，HostA 就可以与HostB 互通了。

更多资料获取：http://learning.huawei.com/cn



如果两个用户属于相同的VLAN，但VLAN内配置了用户隔离。用户间要互通，需要在关联了VLAN的接口上启动VLAN内Proxy ARP功能。

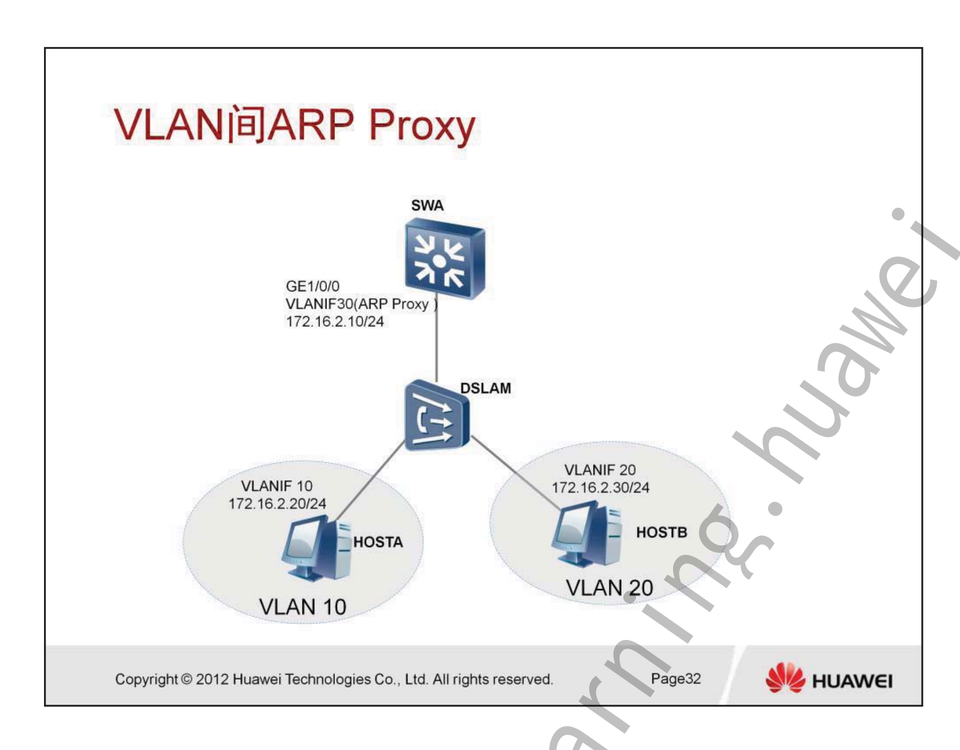
若交换机的接口使能了VLAN内Proxy ARP功能，接口在接收到目的地址不是自己的ARP请求报文后，交换机并不立即丢弃该报文，而是查找该接口的ARP表项。如果满足代理条件，则将交换机的MAC地址发送给ARP请求方。

VLAN内Proxy ARP主要用于配置了用户隔离的VLAN内的用户间互通。

HOST A和HOST B是DSLAM 设备下的两个用户。连接HOST A和HOST B的两个接口在DSLAM 上属于同一个VLAN10。由于在DSLAM上配置了VLAN 内不同接口彼此隔离，因此HOST A和HOST B不能直接在二层互通。

如果在S5700上创建接口VLANIF10。在VLANIF10上使能VLAN 内Proxy ARP，HOST A和HOST B就可以在三层互通了。VLANIF10的IP地址与VLAN10中的主机IP地址必须在同一个网段。

更多资料获取：http://learning.huawei.com/cn



如果两个用户属于不同的VLAN，用户间要进行三层互通，需要在关联了VLAN的接口上启动VLAN间Proxy ARP功能。

若交换机的接口使能了VLAN间Proxy ARP功能，接口在接收到目的地址不是自己的ARP请求报文后，交换机并不立即丢弃该报文，而是查找该接口的ARP表项。如果满足代理条件，则将交换机的MAC地址发送给ARP请求方。

VLAN间Proxy ARP主要用于：

使得处于不同VLAN的用户进行三层通信。

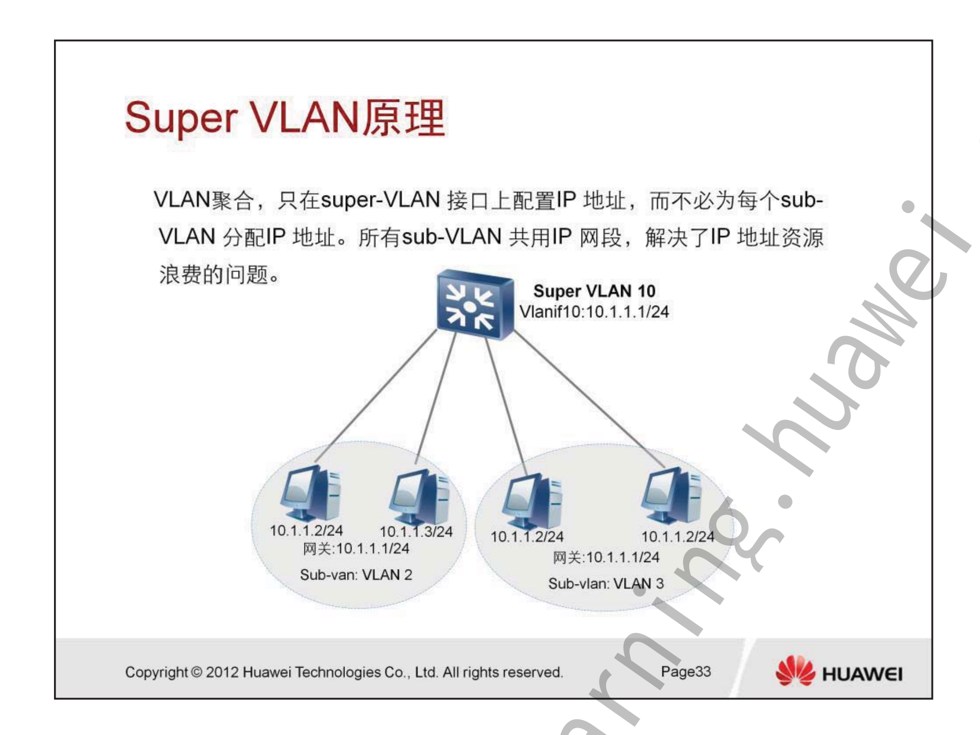
可在Super VLAN对应的VLANIF接口上启动VLAN间Proxy ARP功能，实现Sub VLAN间用户互通。

如图所示：

HOST A和HOST B是DSLAM设备下的两个用户。由于连接HOST A和HOST B的两个接口在DSLAM 上属于不同的VLAN，因此HOST A和HOST B不能直接实现二层互通。

如果在SWA上创建Super VLAN 30，将VLAN10和VLAN20加入VLAN30，并且创建接口VLANIF 30，在VLANIF 30上使能VLAN 间Proxy ARP，HOST A和HOST B就可以实现三层互通了。VLAN IF 30的IP地址与VLAN10和VLAN20中的主机IP地址在同一个网段。

更多资料获取：http://learning.huawei.com/cn



通过VLAN接口实现VLAN间通信时，需要为每个VLAN 的VLAN接口配置一个IP地址。如果VLAN 很多，将占用许多IP地址资源，导致IP地址的浪费。

VLAN aggregation 就是在一个物理网络内，用多个VLAN 隔离广播域，使不同的VLAN属于同一个子网。

用于隔离广播域的VLAN 叫做sub-VLAN，与该子网对应的VLAN叫做super-VLAN。多个sub-VLAN 组成一个super-VLAN。

如图所示，super-VLAN10由sub-VLAN2和sub-VLAN3组成。sub-VLAN2、sub-VLAN3和super-VLAN10属于同一个子网10.1.1.0/24。

super-VLAN10的Vlanif 10的接口地址作为sub-VLAN 2和sub-VLAN 3包含的主机的网关地址。

不同sub-VLAN 下的主机不能互通。如果互通，需要在Super-VLAN 的VLAN 接口上使能ARP Proxy。

http://learning.huawei.com/cn 更多资料获取

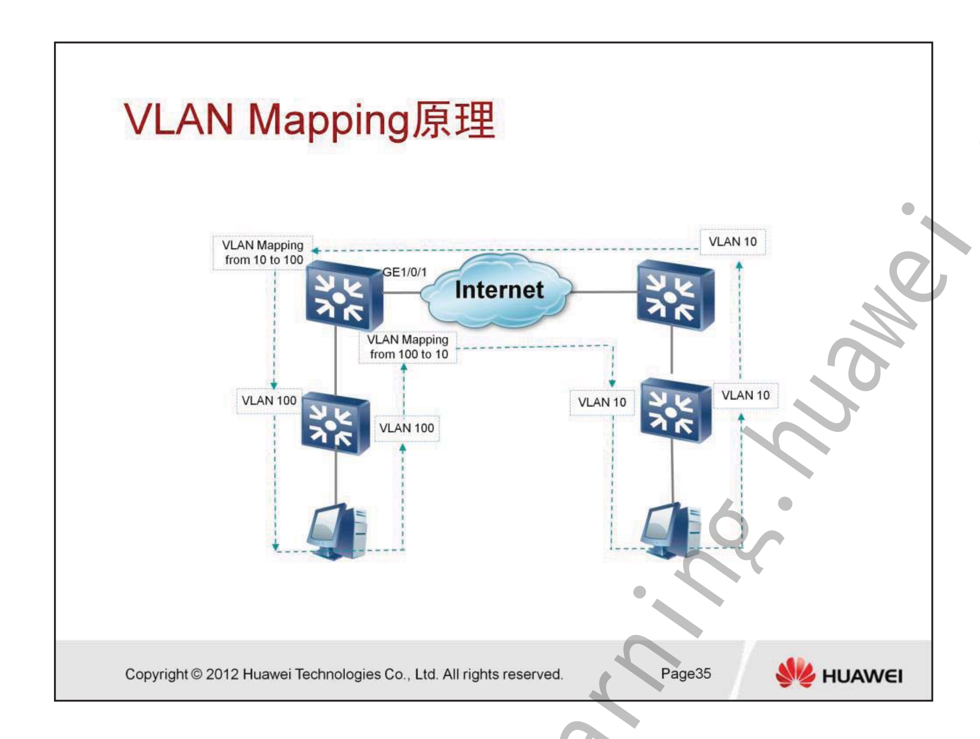
VLAN Mapping概述

VLAN Mapping 也叫做VLAN translation，可以实现在用户VLAN ID（私有VLAN）和运营商VLAN ID(业务VLAN,也可以说是公有VLAN)之间相互转换的一个功能。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn



VLAN Mapping 发生在报文从入端口接收进来之后，从出端口转发出去之前。

当在端口配置了VLAN ID 映射后，端口在向外发送本地VLAN 的帧时，将帧中的VLAN Tag 替换成外部VLAN 的VLAN Tag；在接收外部VLAN 的帧时，将帧中的VLAN Tag替换成本地VLAN的VLAN Tag，这样不同VLAN 间就实现了互相通信。

如图所示，当在端口GE1/0/1 上配置了VLAN100 和VLAN10 映射后，端口在向外发送VLAN100 的帧时，将帧中的VLAN Tag 替换成VLAN10 的VLAN Tag；在接收VLAN10 的帧时，将帧中的VLAN Tag 替换成VLAN100 的VLAN Tag，这样VLAN100 和VLAN10 就实现了互相通信。

更多资料获取：http://learning.huawei.com/cn

VLAN Mapping的配置

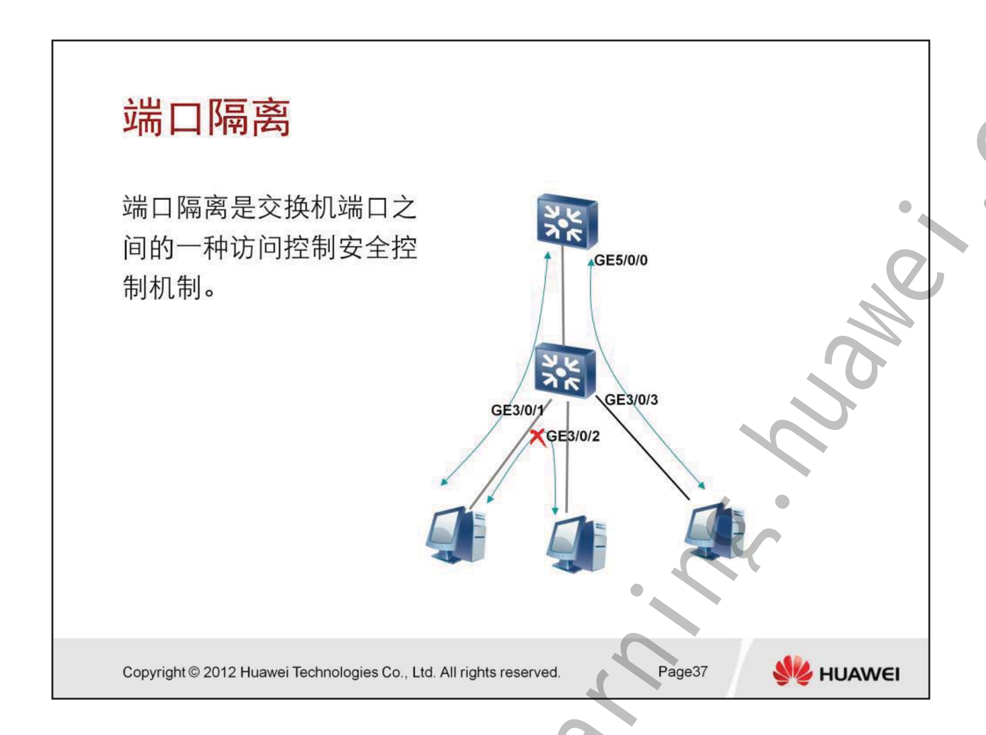
```
[Quidway] interface gigabitethernet 2/0/1
[Quidway-GigabitEthernet2/0/1]port link-type trunk
//配置接口的类型
[Quidway-GigabitEthernet2/0/1]port trunk allow-pass vlan
  100
//配置接口允许通过的VLAN，为mapping后的VLAN
[Quidway-GigabitEthernet2/0/1]qinq vlan-translation enable
//使能接口VLAN转换功能
[Quidway-GigabitEthernet2/0/1]port vlan-mapping vlan 1 to
  20 map-vlan 100
//配置接口2/0/1上VLAN 1~20的报文mapping成VLAN100
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page36 HUAWEI

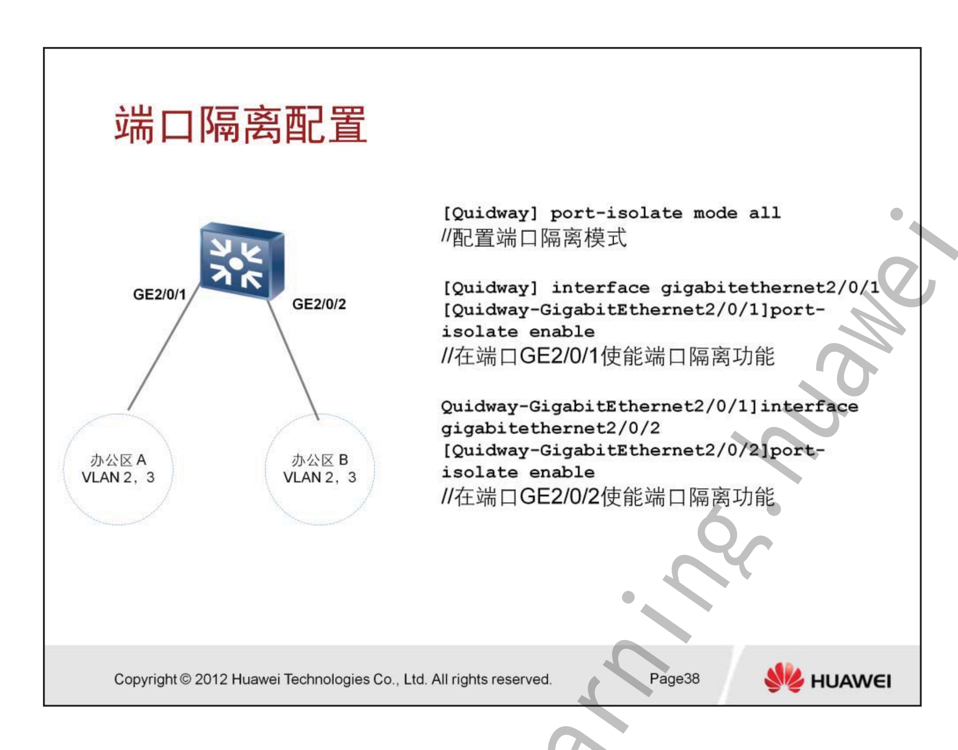
更多资料获取：http://learning.huawei.com/cn





如图要实现不同端口接入的PC之间不能互访（PC属于相同的VLAN），而所有的PC都能够通过上行的交换机访问网络，可以通过配置端口GE3/0/1 、GE3/0/2、GE3/0/3之间端口隔离，GE5/0/0和端口GE3/0/1 、GE3/0/2、GE3/0/3 之间不隔离，来实现此安全控制需求。

更多资料获取：http://learning.huawei.com/cn



如图所示，需要配置办公区A和办公区B相互隔离。办公区A可能有多个VLAN的用户，办公区B也可能有多个VLAN的用户，可以通过配置端口隔离来实现A，B的隔离。

Mux VLAN可以配置VLAN内的用户间的相互隔离或者互通，

端口隔离则是物理层次上的隔离，是基于端口，配置端口隔离后，无论是那个VLAN，两个端口间都不能通信。

更多资料获取：http://learning.huawei.com/cn



问 题

Mux VLAN内分几种类型的VLAN?

什么叫ARP Proxy？

VLAN mapping的作用是什么?

端口隔离的作用是什么?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page39



Q: Mux VLAN内分几种类型的VLAN?

A:Mux VLAN分为主VLAN和从VLAN，从VLAN又分为隔离型从VLAN和互通型从VLAN。

Q:什么叫ARP Proxy?

A: 一个物理网络的子网（Subnet）中的源主机向另一个物理网络的子网中的目的主机发ARP request，和源主机直连的网关用自己接口的MAC地址代替目的主机回ARP reply，这个过程称为ARP代理。

Q:VLAN mapping的作用是什么?

A:VLAN mapping是可以实现在用户VLAN ID（私有VLAN）和运营商VLAN ID(业务VLAN,也可以说是公有VLAN)之间相互转换的一个功能。

Q: 端口隔离的作用是什么?

A: 端口隔离是交换机端口之间的一种访问控制安全控制机制，用于控制两个物理端口的相互访问权限。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

QinQ协议在用户私网VLAN tag之外封装公网VLAN tag，在公网中报文只根据公网VLAN Tag传播。QinQ为用户提供一种较为简单的二层VPN隧道。本文旨在介绍QinQ的基本原理和实现。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 了解QinQ的基本原理和实现方式
- 学会QinQ的简单配置
- 掌握QinQ技术的应用

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page2 HUAWEI

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

QinQ概述

什么是QinQ

- 基于802.1 Q封装的隧道协议
- 报文封装双层VLAN Tag

QinQ优点

- 解决日益紧缺的公网VLAN ID资源问题
- 用户可以规划自己的私网VLAN ID
- 提供一种较为简单的二层VPN解决方案
- 使用户网络具有较高的独立性

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.





QinQ是基于802.1 Q封装的隧道协议，其核心思想是在用户私网VLAN tag之外封装公网VLAN tag，报文带着两层tag穿越公网，从而为用户提供一种较为简单的二层VPN隧道。

QinQ协议是一种简单且易于管理的协议，它不需要信令的支持，仅仅通过静态配置即可实现，特别适用于以三层交换机为骨干的小型企业网或小规模城域网。

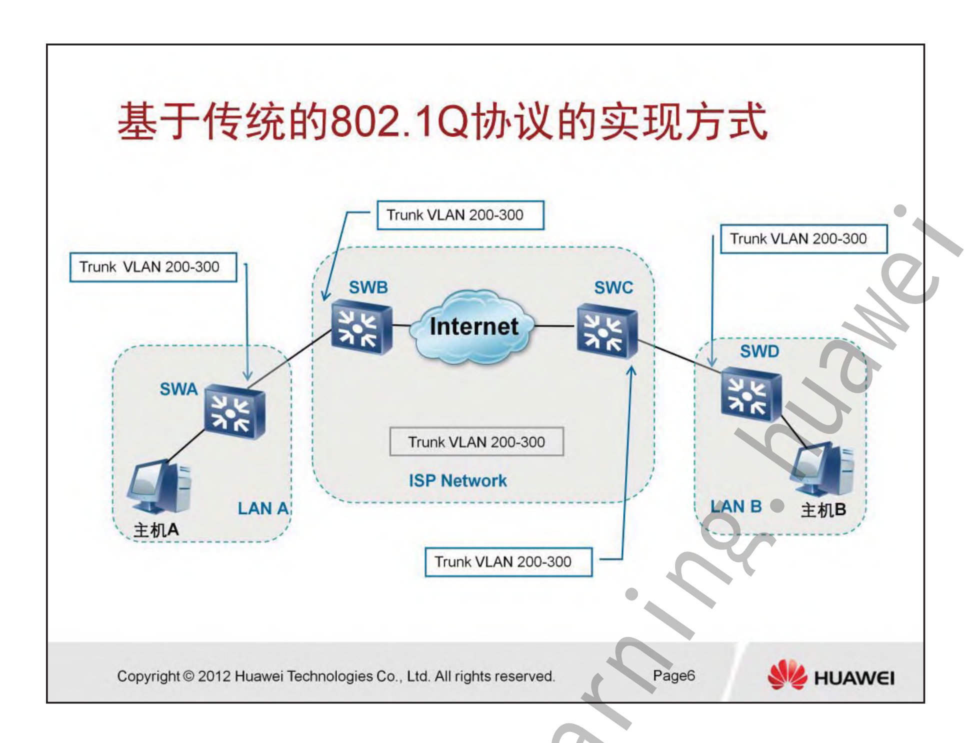
QinQ协议在解决小型城域网或企业网方案时，具有以下优点：

- 可以解决日益紧缺的公网VLAN ID资源问题；
- 用户可以规划自己的私网VLAN ID，不会导致与公网VLAN ID冲突；
- 提供一种较为简单的二层VPN解决方案；
- 使用户网络具有较高的独立性，在服务提供商升级网络时，用户网络不必更改原有的VALN ID配置。

更多资料获取：http://learning.huawei.com/cn

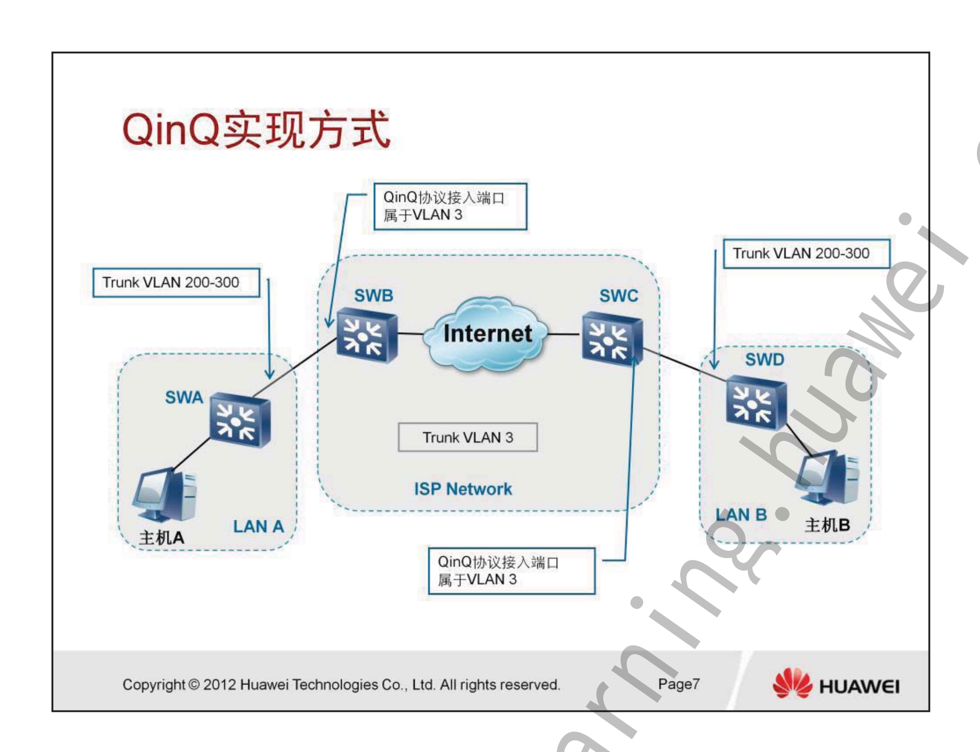


更多资料获取：http://learning.huawei.com/cn

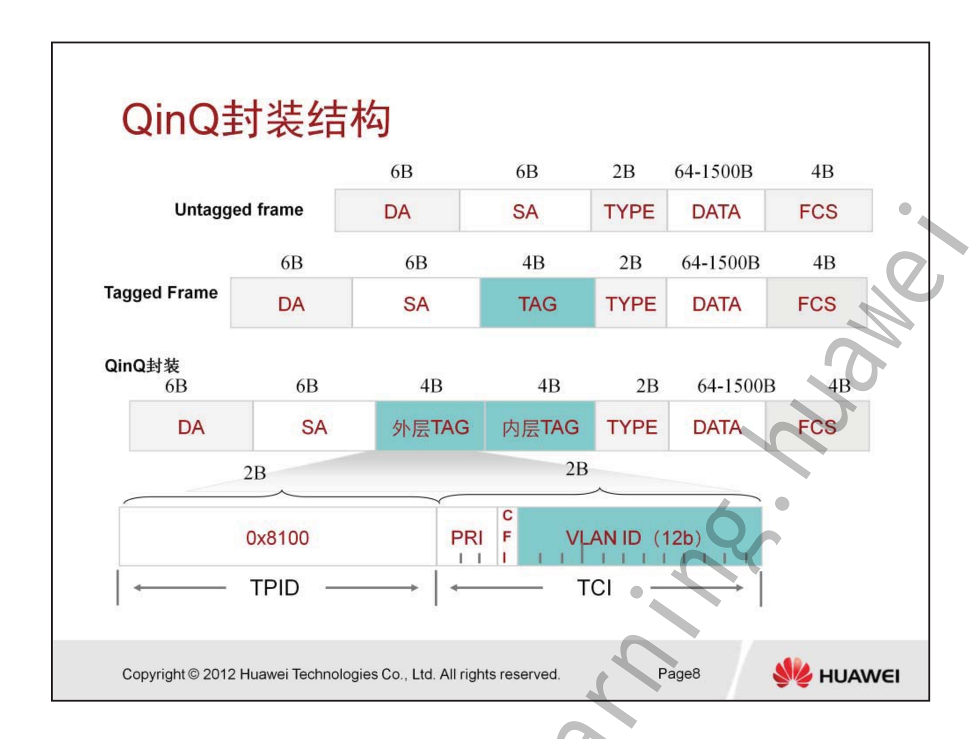


如图所示，假设某企业的LAN A和LAN B位于两个不同地点，分别通过SWA和SWD接入到服务提供商骨干网络（ISP Network），对于传统的802.1Q协议的网络，如果用户需要将LAN A的VLAN200-300和LAN B的VLAN200-300分别互联起来，那么必须将SWA、SWB和SWC、SWD的相连端口都配置为Trunk属性，并允许通过VLAN200-300，这种配置方法必须使用户的VLAN在骨干网络上可见，不仅耗费服务提供商宝贵的VLAN ID资源（一般只有4094个VLAN ID资源），而且还需要服务提供商管理用户的VLAN号，用户没有自己规划VLAN的权利。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



如图所示，普通以太网帧中，没有VLAN Tag，802.1Q帧中在源地址和Type字段之间插入4B长度的VLAN Tag，QinQ封装在802.1Q VLAN Tag之前再插入一个4B长度的VLAN Tag，两个4B长度的VLAN Tag字段包含相同的内容。

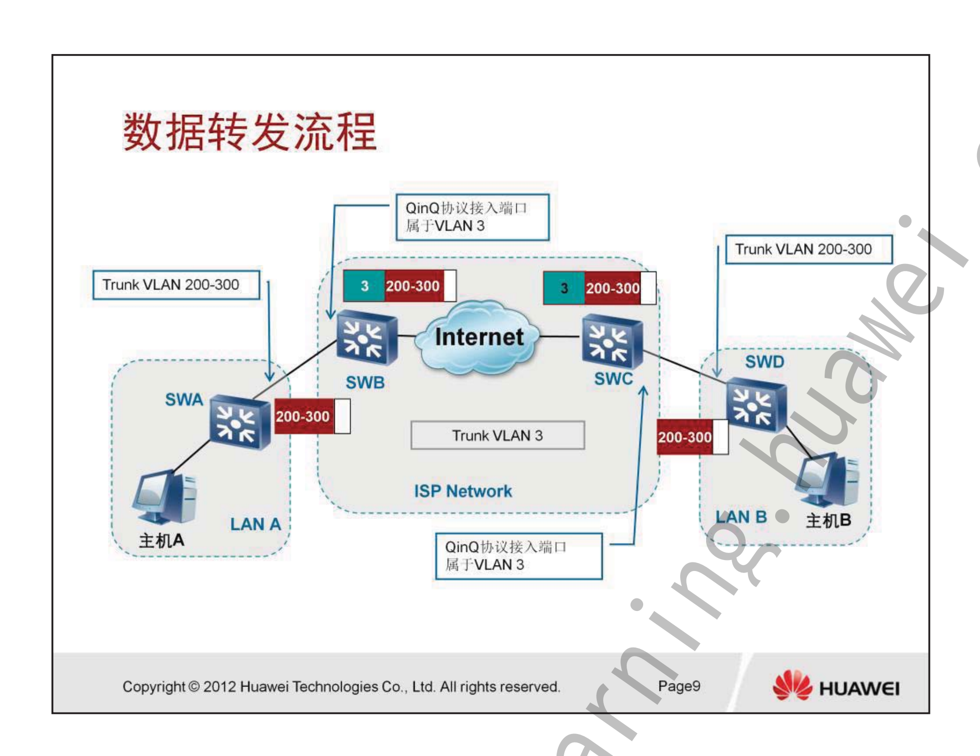
其中：

TPID：Tag Protocol Identifier，2字节，固定取值，0x8100，是IEEE定义的新类型，表明这是一个携带802.1Q标签的帧。华为交换机外层Tag中TPID值缺省采用协议规定的0x8100，某些厂商的设备将QinQ报文外层Tag的TPID值设置为0x9100或0x9200。为了和这些设备兼容，华为交换机提供基于端口的QinQ报文TPID值可调功能。

TCI：Tag Control Information，2字节。帧的控制信息，详细说明如下：

- Priority：3比特，指示以太网帧的优先级。一共有8种优先级，0 – 7，用于提供有差别的转发服务。
- CFI：Canonical Format Indicator，1比特。用于令牌环/源路由FDDI介质访问中指示地址信息的比特次序信息，即先传送的是低特位还是高比特位。
- VLAN Identified：VLAN ID，12比特，取值从0到4095。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

QinQ的分类

根据QinQ的具体实现方式，通常分为如下几类：

- 基于端口的QinQ
  - 基于端口的基本QinQ
- 灵活QinQ
  - VLAN Stacking
- 基于流的灵活QinQ
  - 基于ACL的灵活QinQ

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page10 HUAWEI

更多资料获取：http://learning.huawei.com/cn



基于端口的QinQ

配置了此功能的端口，设备会为从此端口进入的报文打上一层VLAN ID为端口PVID的外层VLAN tag。

基于端口的QinQ通过配置端口类型为dot1q-tunnel实现。

- 当端口类型为dot1q-tunnel时，该端口加入的VLAN不支持二层组播功能。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn

灵活QinQ

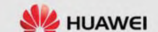
灵活QinQ根据指定条件为入报文加一层S-VLAN tag。

- 指定条件：入报文外层VLAN的范围或VLAN+802.1P。
- 仅指定报文802.1P优先级时，不关注入报文外层VLAN的具体值，只要外层VLAN的802.1P优先级匹配就会打上S-VLAN tag。

通过在端口配置VLAN Stacking实现。

优势:

- 相对基于端口的QinQ，灵活QinQ可以根据入报文的外层VLAN及802.1P来选择加或不加S-VLAN tag，并且S-VLAN tag可配置。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page12 

配置灵活QinQ时，需要配置端口加入S-VLAN的方式为UNTAGGED方式，报文下行时就能够剥掉S-VLAN tag后再出设备。

- 接口类型为Hybrid或Trunk时，才能配置灵活QinQ。
- Trunk类型端口只有在允许通过的VLAN与端口的PVID相同时，端口才以UNTAGGED方式加入VLAN，所以Trunk类型端口只能支持配置一条灵活QinQ。

端口学习MAC地址时，学习的是QinQ报文外层VLAN的MAC地址。

更多资料获取：http://learning.huawei.com/cn

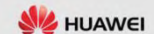
基于流的灵活QinQ

基于流的灵活QinQ通过全局配置流分类、流行为，再将流策略绑定流分类和流行为来实现。

优势:

- 相对灵活QinQ，基于流的灵活QinQ还可以根据入报文的内层VLAN的属性来加S-VLAN tag，配置范围更加灵活。如:
  - 内层VLAN、内层VLAN+802.1P、外层VLAN、外层VLAN+802.1P等属性

了解基于流的灵活QinQ需要先掌握QoS知识，在本课程中仅需了解基本概念。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page13 

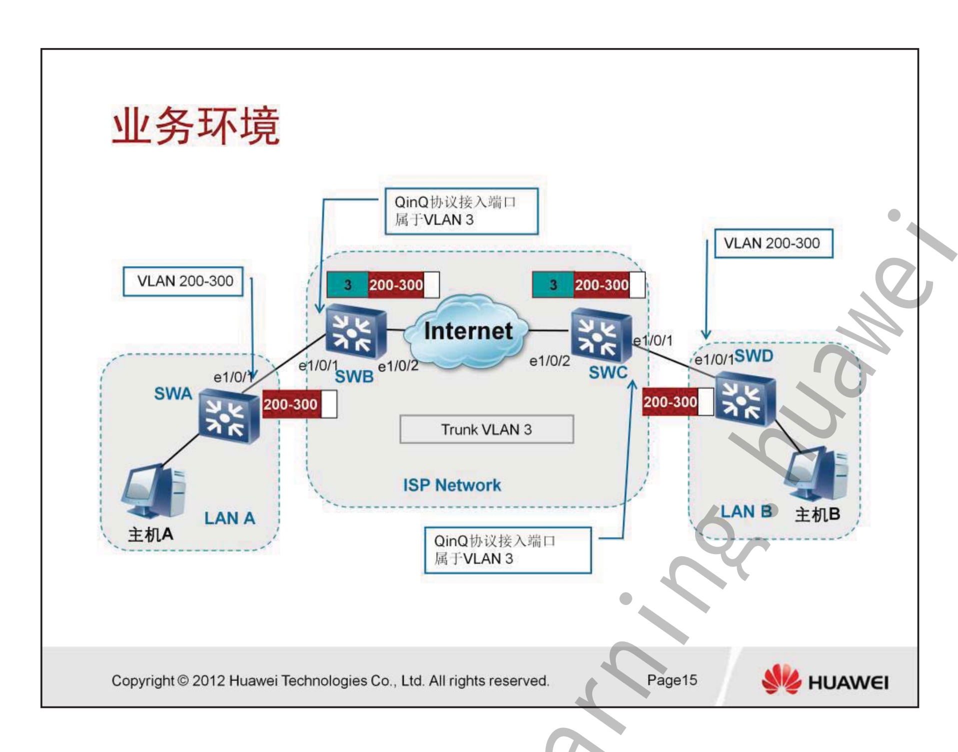
流策略可以在接口视图、VLAN视图和全局视图下配置，不同视图下的配置决定了流策略的不同作用范围。

当流策略配置在VLAN视图，并且流分类指定外层VLAN或者外层VLAN+802.1P时，只有入报文的外层VLAN ID与所配置的VLAN视图对应的VLAN ID相同时，流策略才会生效。

更多资料获取：http://learning.huawei.com/cn



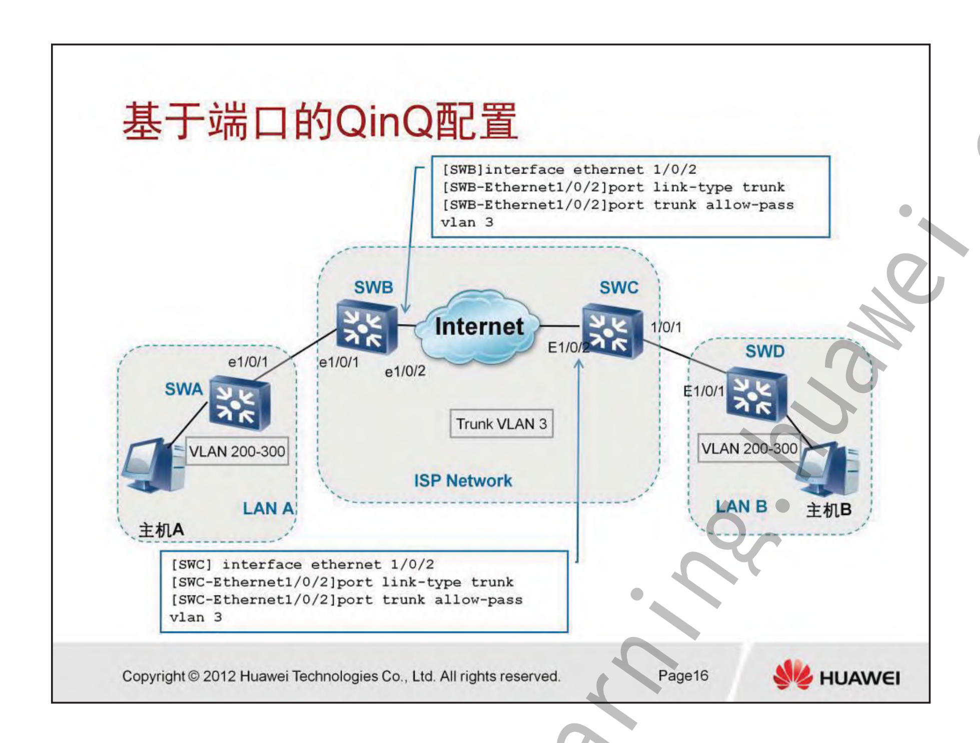
更多资料获取：http://learning.huawei.com/cn



以此组网为例，介绍QinQ基本配置：

- 如图所示，LAN A、LAN B中SWA和SWD连接有VLAN200到VLAN300的私网用户。
- SWB连接SWA的接口Ethernet1/0/1、SWC连接SWD的接口Ethernet1/0/1为QinQ接入端口，属于VLAN3。
- SWA和SWD为用户侧交换机，上行Trunk端口Ethernet1/0/1将带有VLAN ID 200-300的帧发送给SWB和SWC。

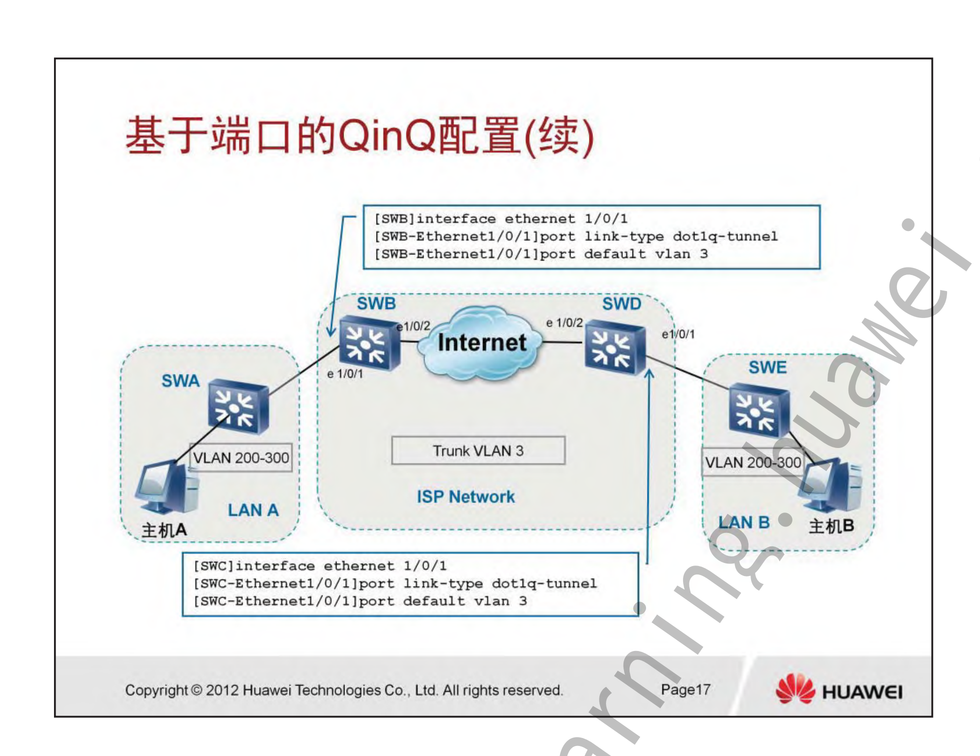
更多资料获取：http://learning.huawei.com/cn



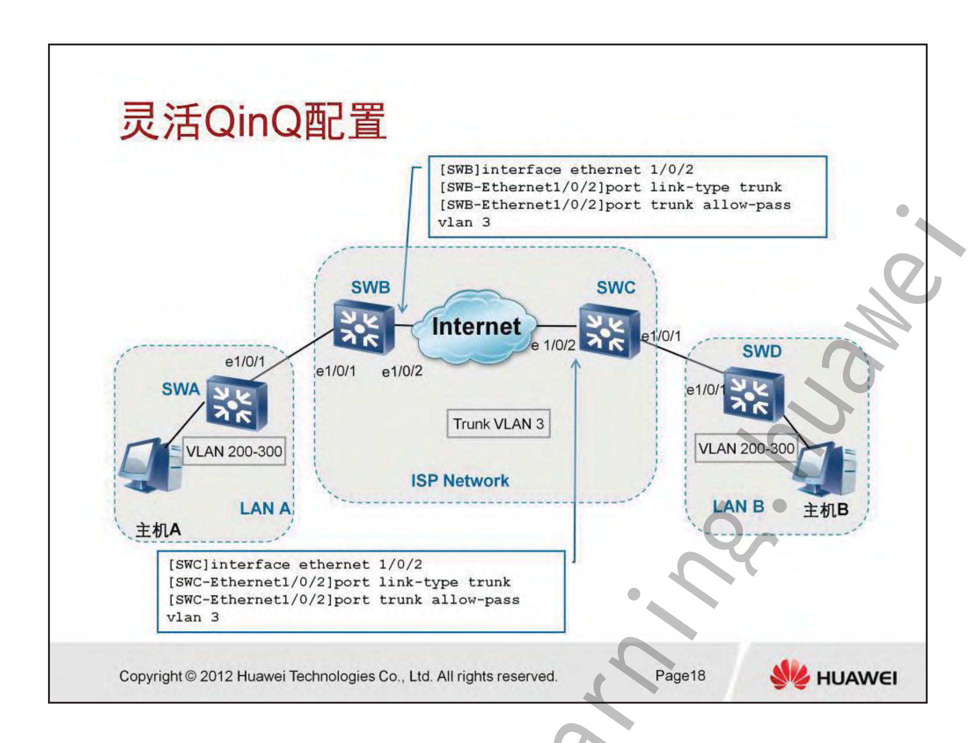
基于端口的QinQ基本配置：

- SWA上行Trunk端口具体配置如下：
  - [SWA]interface ethernet1/0/1
  - //进入ethernet 1/0/1接口模式。
  - [SWA-Ethernet1/0/1]port link-type trunk
  - //配置链路类型为trunk链路。
  - [SWA-Ethernet1/0/1]port trunk allow-pass vlan 200 to 300
  - //配置允许带有VLAN ID 200到300的帧穿过。
- SWD上行Trunk端口配置与SWA类似。

更多资料获取：http://learning.huawei.com/cn



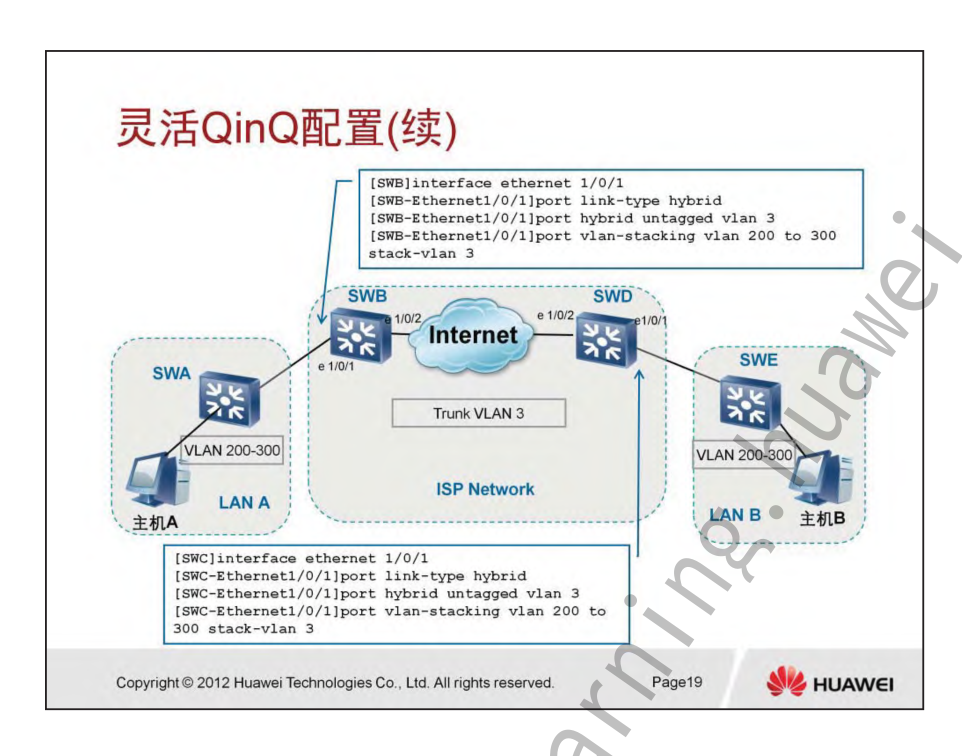
更多资料获取：http://learning.huawei.com/cn



灵活QinQ基本配置:

- SWA上行Trunk端口具体配置如下:
  - [SWA]interface ethernet1/0/1
  - //进入ethernet 1/0/1接口模式。
  - [SWA-Ethernet1/0/1]port link-type trunk
  - //配置链路类型为trunk链路。
  - [SWA-Ethernet1/0/1]port trunk allow-pass vlan 200 to 300
  - //配置允许带有VLAN ID 200到300的帧穿过。
- SWD上行Trunk端口配置与SWA类似。

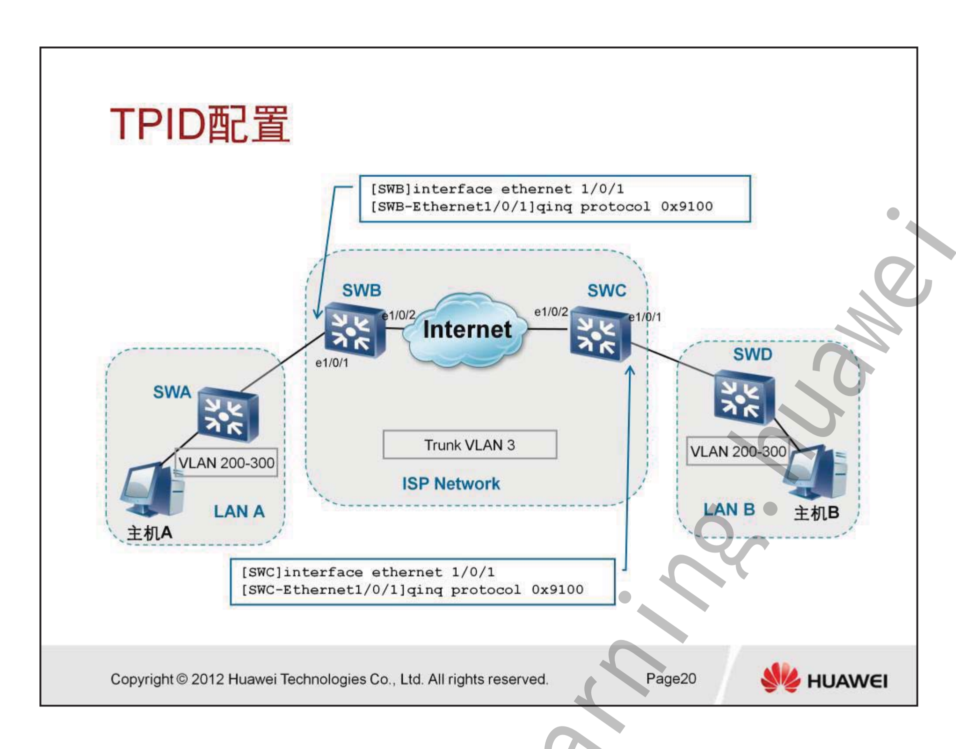
更多资料获取：http://learning.huawei.com/cn



配置命令说明:

- port link-type hybrid
  - 灵活QinQ的接入端口类型可以配置为trunk或hybrid，根据具体情况而定。默认为hybrid，此时这条命令可不配。
- port hybrid untagged vlan 3
  - 配置接口以Untagged方式加入叠加后的VLAN。
  - 叠加后的外层VLAN必须是设备上已存在的VLAN，叠加前的VLAN可以不创建。
  - 配置后，接口在发送帧时会将帧中的Tag剥离。
  - 如果在同一接口上多次使用此命令，则最终是多次配置的合集。
- port vlan-stacking vlan 200 to 300 stack-vlan 3
  - 配置灵活QinQ，在VLAN 200到300的用户VLAN上添加外层VLAN 3。

更多资料获取：http://learning.huawei.com/cn



默认为8100，可配置范围为0x0600～0xFFFF。

更多资料获取：http://learning.huawei.com/cn



问 题

QinQ的作用是什么？如何实现?

QinQ报文在公网上传递是否会检查内层私网Tag?

QinQ报文公网Tag中TPID缺省值是什么？是否可以修改?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page21



Q:QinQ的作用是什么?

A:QinQ为用户提供一种较为简单的二层VPN隧道，通过在用户私网VLAN tag之外封装公网VLAN tag实现。

Q:QinQ报文在公网上传递是否会检查内层私网Tag?

A:QinQ报文只按照外层公网Tag在公网上传递，不会检查内层私网Tag。

Q:QinQ报文公网Tag中TPID缺省值是什么？是否可以修改?

A:QinQ报文公网Tag中TPID缺省值是0x8100，可以使用命令qinq protocol修改。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn

Module 2

STP

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

本课程介绍STP（生成树协议）的原理与配置。

STP运行于以太网交换机上，通过在网络上修剪出一棵无环的树来解决交换网络中的环路问题。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1 

本课程介绍STP（生成树协议）的原理与配置。

STP：Spanning Tree Protocol。

以太网交换网络上为了进行链路备份，通常会使用冗余链路，但是使用冗余链路会在交换网络上生成环路，并导致广播风暴以及MAC地址表不稳定等故障现象。

STP运行于以太网交换机上，为解决交换网络中的环路问题在网络上修剪出一棵无环的树，并在主链路故障后，自动启用备份链路，使网络工作正常。

最新的STP标准由1998年发布的IEEE802.1D标准文档定义。

更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 描述生成树基本计算过程
- 描述配置BPDU在计算过程中的作用
- 描述拓扑结构改变信息的泛洪过程

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page2



学习完此课程，您将会：

描述生成树基本计算过程；

描述配置BPDU在计算过程中的作用，理解交换机如何选择最优的配置BPDU，以及如何设置端口状态；

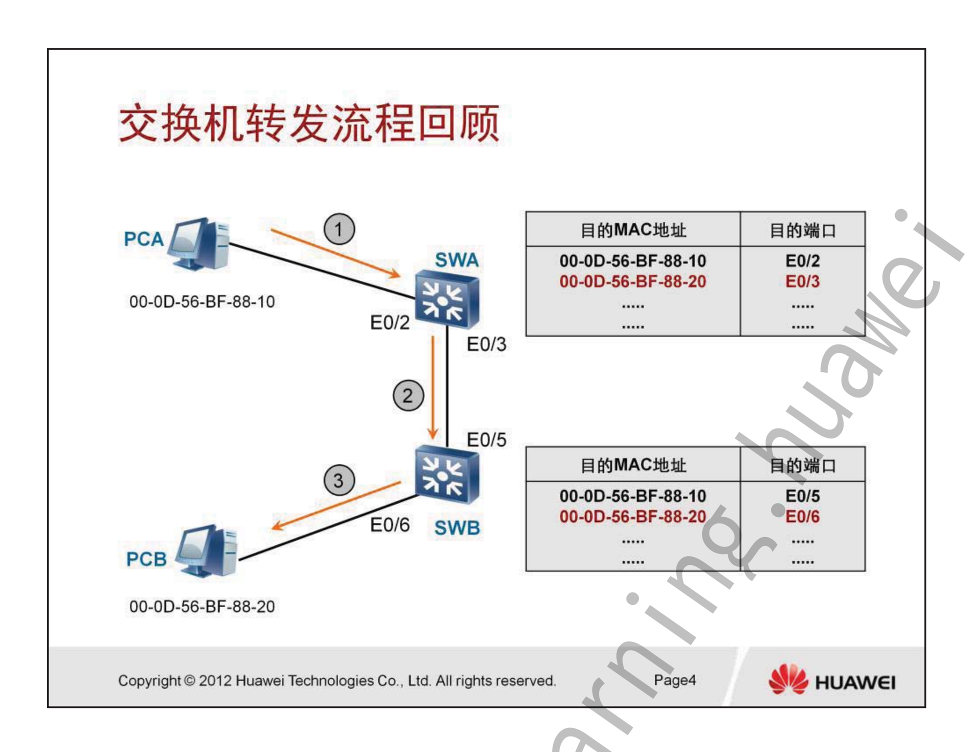
描述拓扑结构改变信息的泛洪过程，理解拓扑结构改变信息的作用。

更多资料获取：http://learning.huawei.com/cn



本章通过回顾交换机的工作过程，了解交换网络中的环路引起的问题，理解为什么要使用STP（生成树协议）。

更多资料获取：http://learning.huawei.com/cn



交换机基于MAC地址表进行转发数据帧，MAC地址表是目的MAC地址和目的端口的对应关系。

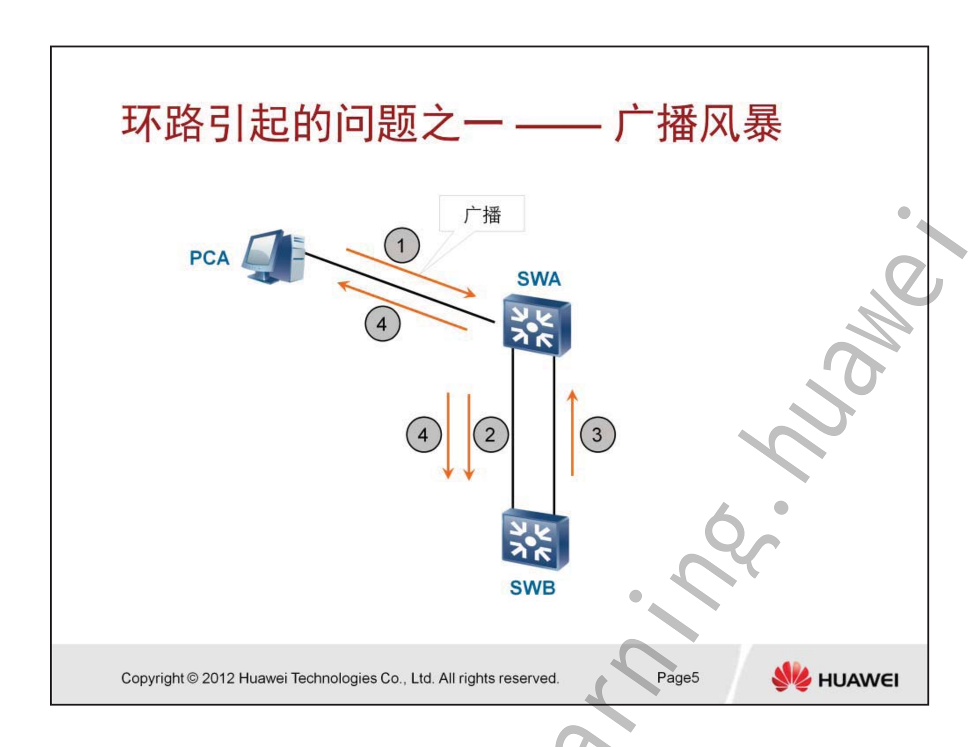
1：假设PCA向PCB发送一个数据帧，此数据帧的目的MAC地址设置为PCB的MAC地址00-0D-56-BF-88-20，交换机SWA接收到此数据帧之后，需要查找MAC地址表，根据MAC地址表中的记录，将数据帧从E0/3口向外转发。

交换机在转发数据帧的时候，对数据帧不做任何修改，如果交换机接收到的是一个广播数据帧，则向除源端口之外的所有端口转发。

2：交换机SWB接收到了此数据帧之后，查找MAC地址表，根据MAC地址表中的记录，将数据从E0/6端口上转发出去，此次转发仍然不会对数据帧做任何修改。

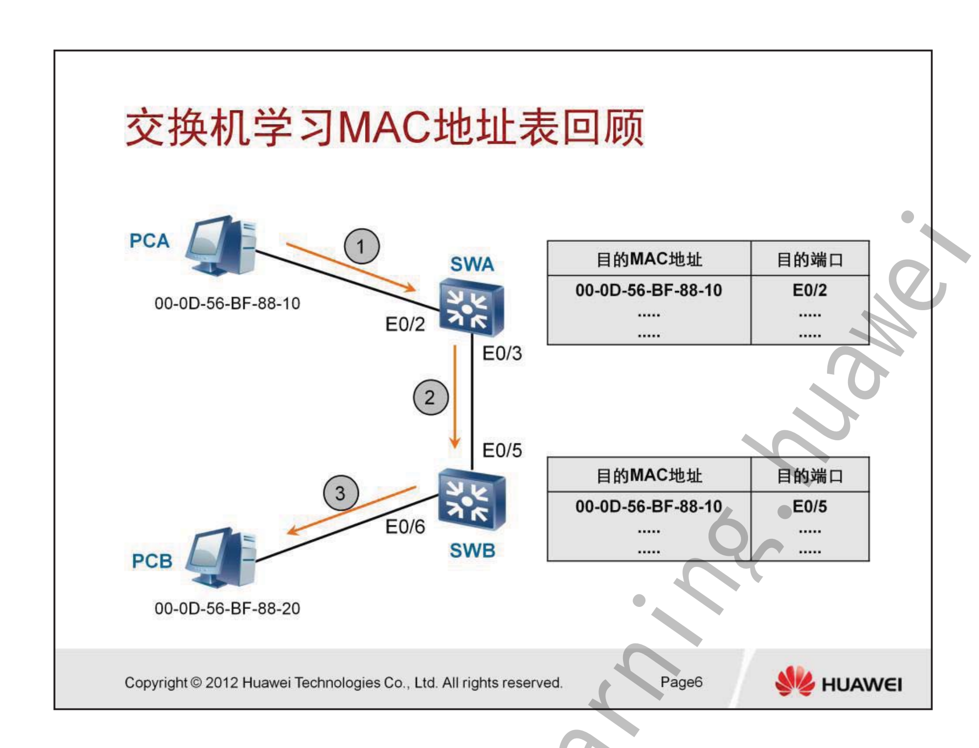
3：PCB接收到数据帧之后，查看目的MAC地址，由于目的MAC地址为接收者本身，所以PCB处理此数据帧并上送给上层协议处理数据帧所携带的数据。

更多资料获取：http://learning.huawei.com/cn



如果交换机从一个端口上接收到的是一个广播数据帧，则向所有其它端口转发，而且交换机在转发数据帧的时候，对数据帧不做任何修改，因此，如果交换网络中有环路，则广播帧会被无限期的转发，形成广播风暴。

更多资料获取：http://learning.huawei.com/cn



交换机根据MAC地址表转发，但是MAC地址表在交换机启动时是空的，交换机有一个学习MAC地址表的过程。

交换机是根据接收到的数据帧的源地址和接收端口的对应关系学习MAC地址表的。

1：假设PCA向PCB发送一个数据帧。在此数据帧中，目的MAC地址是PCB的MAC地址00-0D-56-BF-88-20，源地址是PCA的MAC地址00-0D-56-BF-88-10。

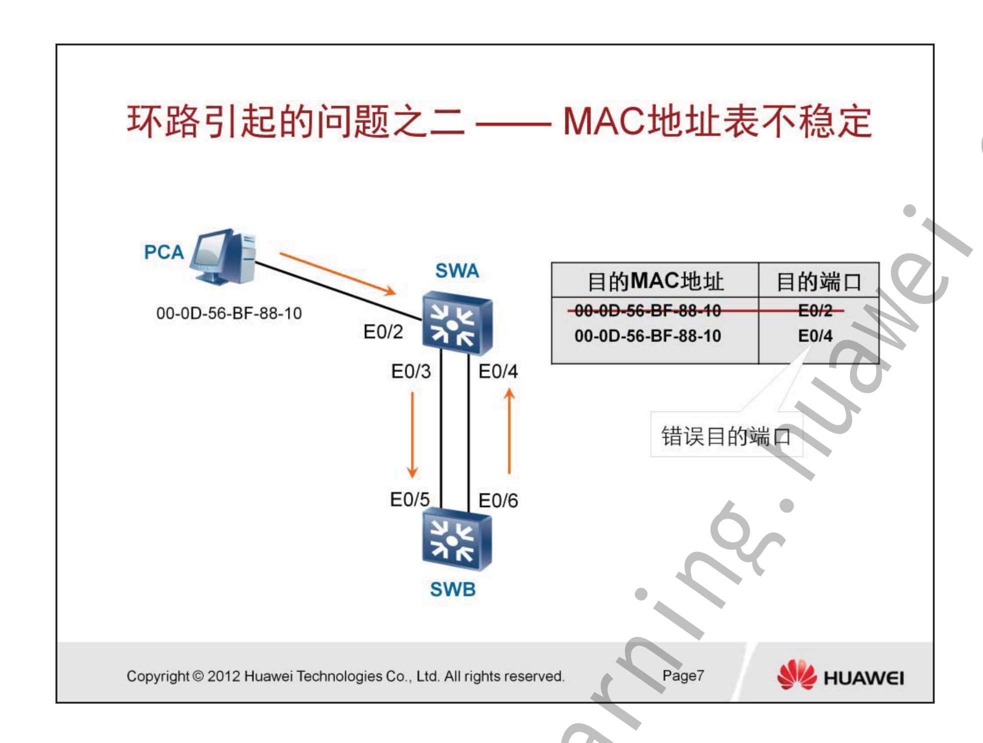
当交换机SWA收到此数据帧之后，检查数据帧的源地址，并将源地址和接收端口的对应关系添加到MAC地址表中，形成目的地址和目的端口的对应关系。

2：交换机SWB收到此数据帧之后，同样将源MAC地址和接收端口的对应关系添加到MAC地址表中，形成一个MAC地址表项。

3：PCB收到数据帧之后，处理数据帧。



更多资料获取：http://learning.huawei.com/cn



交换机根据所接收到的数据帧的源地址和接收端口的对应关系生成MAC地址表。

PCA向外发送一个数据帧，假设此数据帧的目的MAC地址在网络中所有交换机的MAC地址表中都暂时不存在。SWA收到此数据帧之后，在MAC地址表中生成一个MAC地址表项，00-0D-56-BF-88-10，对应端口为E0/2。

由于SWA的MAC地址表中没有对应此数据帧目的MAC地址的表项，则SWA将此数据帧同时向E0/3和E0/4端口上转发。

由于SWB的MAC地址表中也没有对应此数据帧目的MAC地址的表项，则从E0/5接口接收到的数据帧会被从E0/6接口发送回SWA。

SWA从E0/4接收到此数据帧之后，会在MAC地址表中删除原有的相关表项，生成一个新的表项，00-0D-56-BF-88-10，对应端口为E0/4。这样不但造成MAC地址表不稳定，而且还生成了错误的表项。

更多资料获取：http://learning.huawei.com/cn



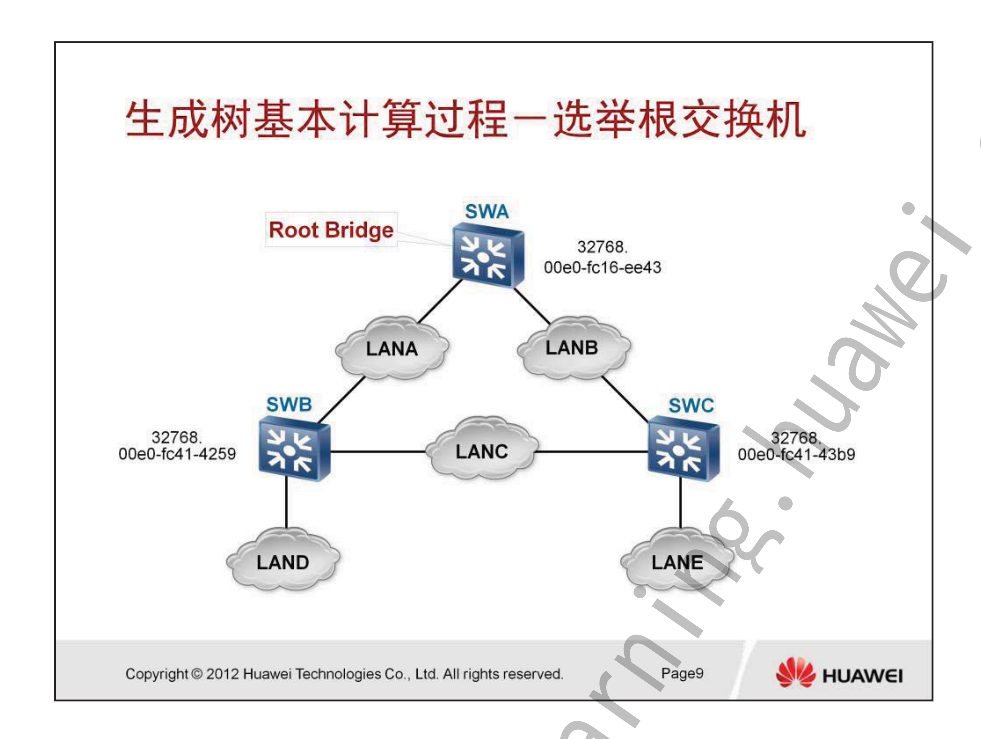
目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page8



本章介绍生成树计算的基本过程，理解生成树协议中的基本概念，包括交换机角色，端口角色，端口状态等内容。

更多资料获取：http://learning.huawei.com/cn



为了计算生成树，交换机之间需要交换相关信息和参数，这些信息和参数被封装在配置BPDU（Configuration Bridge Protocol Data Unit）中，在交换机之间传递。

BPDU是指桥接协议数据单元，泛指交换机之间运行的协议交互信息时使用的数据单元。配置BPDU是BPDU的一种。

生成树计算的第一步是选举根交换机，根交换机的选举基于交换机标识（Bridge ID）。

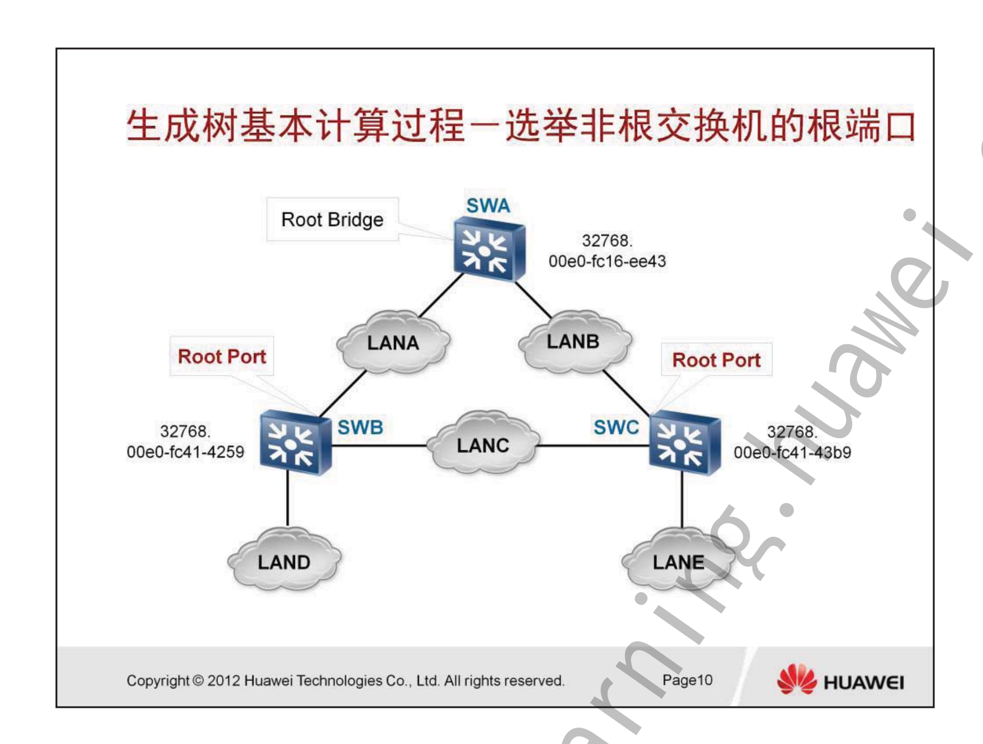
交换机标识由两部分组成：两字节长度的交换机优先级和六字节长度的MAC地址。

交换机优先级是可以配置的，取值范围是0～65535，默认值为32768。

网络中交换机标识最小的成为根交换机，首先比较优先级，如果优先级相同则比较MAC地址，值越小越优先。

本例中，三个交换机的优先级是相同的，由于SWA的MAC地址值最小，因此SWA为根交换机。

更多资料获取：http://learning.huawei.com/cn

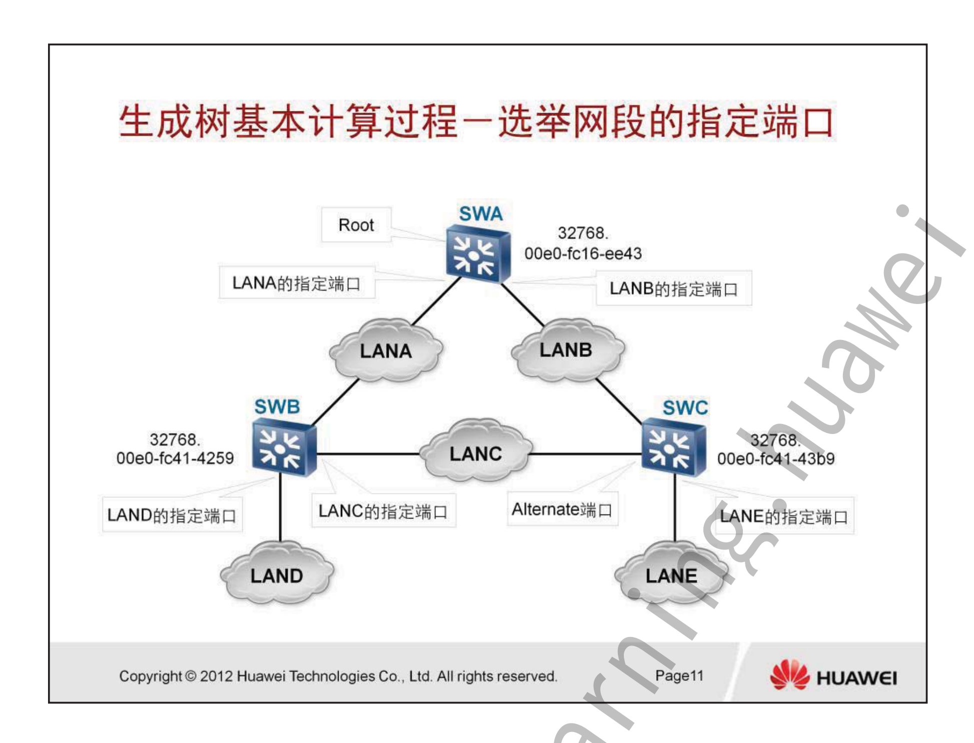


STP为每个非根交换机选举根端口（Root Port）。

交换机的每个端口都有一个端口开销（Port Cost）的参数，此参数表示数据从该端口发送时的开销值，也即出端口的开销。STP认为从一个端口接收数据是没有开销的。端口的开销和端口的带宽有关，带宽越高，开销越小，VRP平台中，百兆端口的开销值为200。从一个非根交换机到达根交换机的路径可能有多条，每一条路径都有一个总的开销值，此开销值是该路径上所有出端口的端口开销总和。

根端口是指从一个非根交换机到根交换机总开销最小的路径所经过的本地端口。这个最小的总开销值称为交换机的根路径开销（Root Path Cost）。如果这样的端口有多个，则比较端口上所连接的上行交换机的交换机标识，越小越优先，如果端口上所连接的上行交换机的交换机标识相同，则比较端口上所连接的上行端口的端口标识（Port Identifier），越小越优先。端口标识由两部分组成：一字节长度的端口优先级和一字节长度的端口号。一字节长度的端口优先级是可配置的，默认为128。本例中，假设所有端口都是百兆端口，使用相同的开销值200。

更多资料获取：http://learning.huawei.com/cn



STP为每个网段选出一个指定端口（Designated Port），指定端口为每个网段转发发往根交换机方向的数据，并且转发由根交换机方向发往该网段的数据。指定端口所在的交换机称为该网段的指定交换机。

为每个网段选举指定端口和指定交换机的时候，首先比较该网段所连接的端口所属交换机的根路径开销，越小越优先；如果根路径开销相同，则比较所连接的端口所属交换机的交换机标识，越小越优先；如果根路径开销相同，交换机标识也相同，则比较所连接的端口的端口标识，越小越优先。

对于根交换机来说，所有端口都是所连网段的指定端口。因此LANA和LANB的指定端口都在SWA上。

LAND和LANE都只连接了一个交换机端口，此端口即为指定端口。

对于LANC来说，同时连接到两个交换机端口，并且两个交换机的根路径开销相同，因此需要比较两个端口所在交换机的交换机标识，由于SWB的交换机标识比SWC小（二者交换机优先级一致，但SWB的MAC地址更小），因此LANC的指定端口在SWB上。

既不是根端口也不是指定端口的交换机端口称为Alternate Port（预备端口），预备端口不转发数据，处于阻塞状态。

更多资料获取：http://learning.huawei.com/cn

交换机端口角色

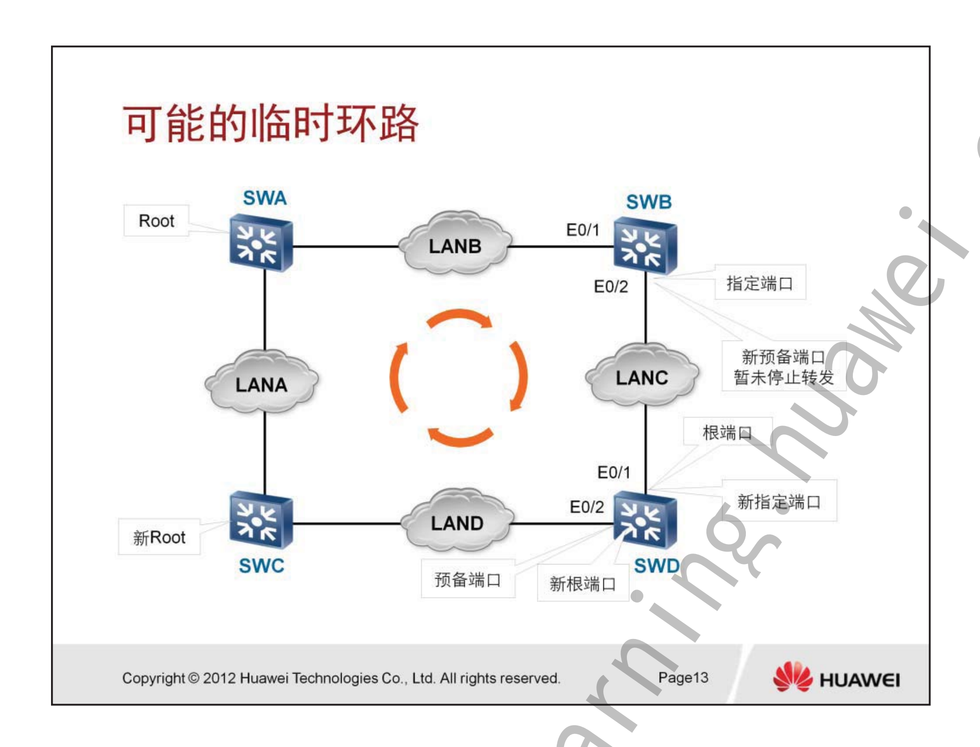
端口角色	描　　述
Root Port	根端口，是所在交换机上离根交换机最近的端口，处于转发状态。
Designated Port	指定端口，转发所连接的网段发往根交换机方向的数据和从根交换机方向发往所连接的网段的数据。
Alternate Port	预备端口，不向所连网段转发任何数据。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page12 

如前所述，对于物理层和数据链路层可以正常工作，并且开启了STP的交换机端口，STP共定义了三种端口角色，处于转发状态的有根端口和指定端口。

底层没有开启的端口称为Disable端口。

更多资料获取：http://learning.huawei.com/cn



在端口角色以及状态的变化过程中，可能会出现临时环路问题。

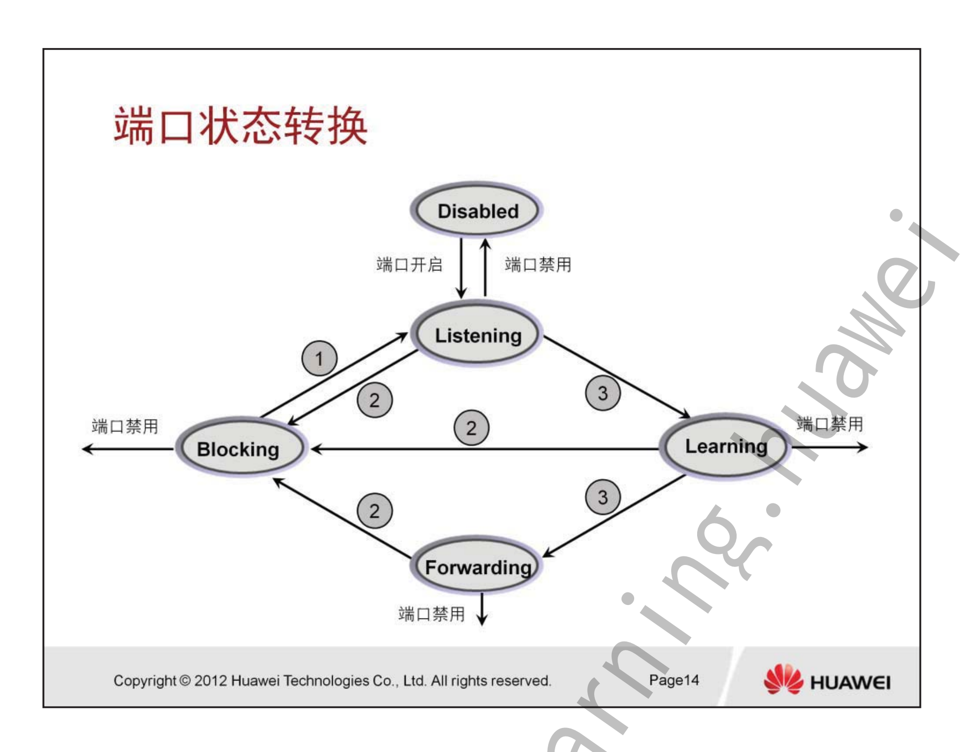
本例中，初始状态下SWA为根交换机，所有的交换机端口中，只有SWD的E0/2端口为Alternate Port，处于不转发状态。

假设修改SWC的优先级，使SWC成为新的根交换机，SWD的E0/2接口成为新的根端口，进入转发状态，E0/1接口成为新的指定端口，处于转发状态，SWB的E0/2应当成为新的Alternate Port，进入不转发状态。

如果在SWB的E0/2在从转发状态进入不转发状态之前，SWD的E0/2就已经从不转发状态进入了转发状态，则网络中会出现临时环路。

解决临时环路的方法是：在一个端口从不转发状态进入转发状态之前（例如SWD的E0/2端口），需要等待一个足够长的时间，以使需要进入不转发状态的端口有足够时间完成生成树计算，并进入不转发状态。

更多资料获取：http://learning.huawei.com/cn



1：端口被选为指定端口（Designated Port）或根端口（Root Port）；

2：端口被选为预备端口（Alternate Port）；

3：经过Forward Delay间隔。Forward Delay默认为15秒。

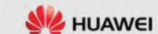
端口被禁用之后进入Disable状态。

当一个端口从不转发状态进入转发状态之前需要等待两次Forward Delay间隔（后文详细解释端口状态变换），以解决前文所述可能的临时环路问题。

更多资料获取：http://learning.huawei.com/cn

端口状态描述

端口状态	描　述
Disable 未启用	此状态下端口不转发数据帧，不学习MAC地址表，不参与生成树计算。
Blocking 阻塞状态	此状态下端口不转发数据帧，不学习MAC地址表，此状态下端口接收并处理BPDU，但是不向外发送BPDU。
Listening 侦听状态	此状态下端口不转发数据帧，不学习MAC地址表，只参与生成树计算，接收并发送BPDU。
Learning 学习状态	此状态下端口不转发数据帧，但是学习MAC地址表，参与计算生成树，接收并发送BPDU。
Forwarding 转发状态	此状态下端口正常转发数据帧，学习MAC地址表，参与计算生成树，接收并发送BPDU。

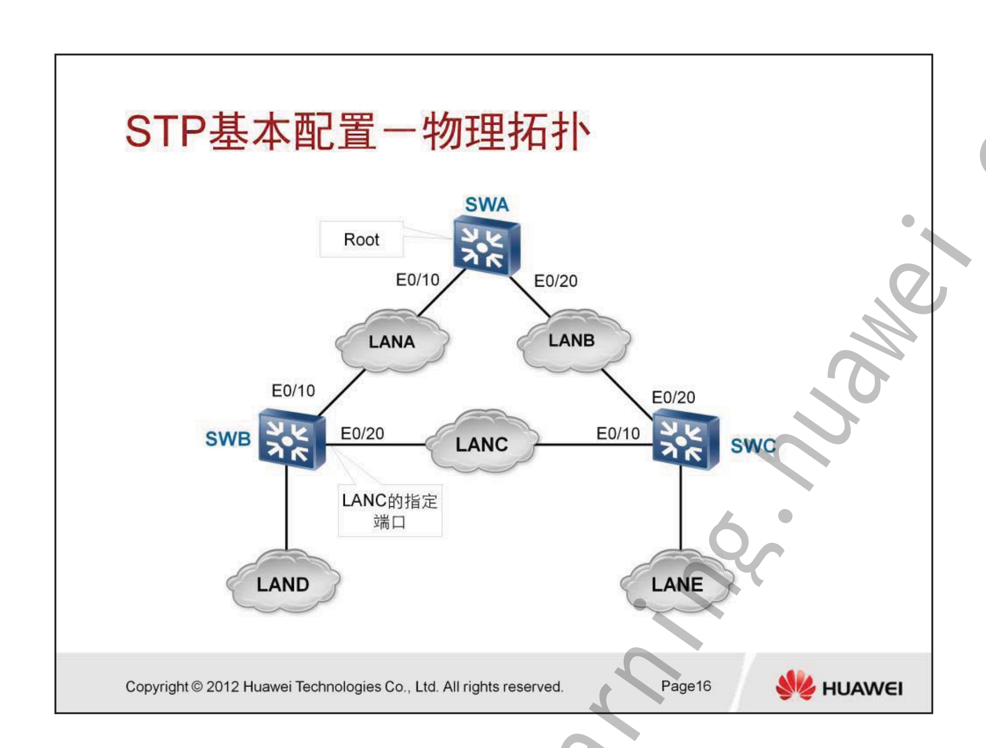
Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page15 

当端口正常启用之后，端口首先进入Listening状态，开始生成树的计算过程。

如果经过计算，端口角色需要设置为预备端口（Alternate Port），则端口状态立即进入Blocking；

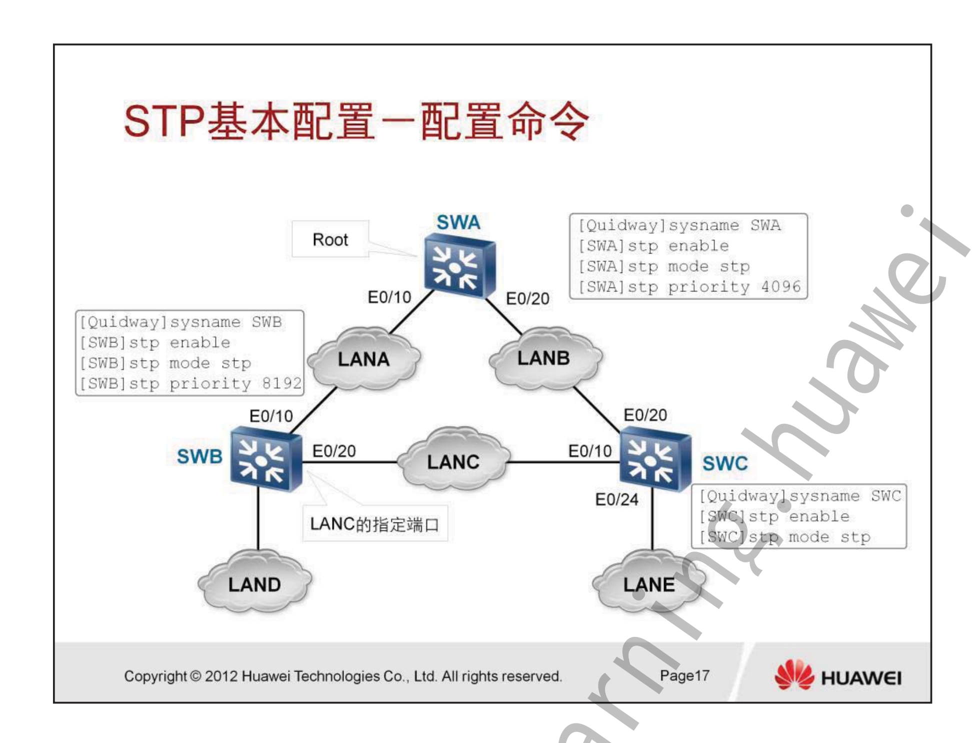
如果经过计算，端口角色需要设置为根端口（Root Port）或指定端口（Designated Port），则端口状态在等待Forward Delay之后从Listening状态进入Learning状态，然后继续等待Forward Delay之后，从Learning状态进入Forwarding状态，正常转发数据帧。

更多资料获取：http://learning.huawei.com/cn



物理拓扑如图所示，配置SWA的Priority为4096、SWB的Priority为8192、SWC的Priority为32678，使SWA成为根交换机，SWB成为LANC的指定交换机。

更多资料获取：http://learning.huawei.com/cn



stp { enable | disable }

stp命令用来启动或关闭交换机全局或端口的STP功能，缺省情况下，交换机上的STP功能处于开启状态。

stp mode { stp | rstp | mstp }

stp mode命令用来设定交换机的STP运行模式，缺省情况下，交换机的运行模式为MSTP模式。

关于RTSP和MSTP技术，将在后续课程中介绍，本课程只介绍STP技术。

stp priority priority

priority：交换机的优先级，取值0～61440，步长为4096，即交换机可以设置16个优先级取值，如0、4096、8192等。

stp priority命令用来配置交换机的优先级，缺省情况下，交换机优先级取值为32768。

更多资料获取：http://learning.huawei.com/cn

STP基本配置－验证STP全局状态

```
[SWC]display stp
-------[CIST Global Info][Mode STP]-------
CIST Bridge          : 32768.00e0-fc41-43b9
Bridge Times         :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC       : 4096.00e0-fc41-4259 / 20000
CIST RegRoot/IRPC    : 32768.00e0-fc41-43b9 / 0
CIST RootPortId      :128.10
BPDU-Protection      :Disabled
TC or TCN received   :117
TC count per hello   :1
STP Converge Mode    :Normal
Share region-configuration :Enabled
Time since last TC   :0 days 0h:0m:0s
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page18 HUAWEI

全局信息中根交换机和自身的交换机标识不同，标识自身是一个非根交换机。

更多资料获取：http://learning.huawei.com/cn

STP基本配置－验证STP端口信息

```
[SWC]display stp interface Ethernet 0/20
----[CIST][Port20(Ethernet0/20)][Forwarding]----
 Port Protocol         :Enabled
 Port Role             :Root Port
 Port Priority         :128
 Port Cost(Dot1T )     :Config=auto / Active=200000000
 Designated Bridge/Port   : 4096.00e0-fc41-4259 / 128.20
 Port Edged            :Config=default / Active=disabled
 Point-to-point        :Config=auto / Active=false
 Transit Limit         :147 packets/hello-time
 Protection Type       :None
 Port STP Mode         :STP
 Port Protocol Type    :Config=auto / Active=dot1s
 BPDU Encapsulation    :Config=stp / Active=stp
 PortTimes             :Hello 2s MaxAge 20s FwDly 15s RemHop 20
 TC or TCN send        :0
 TC or TCN received    :0
 BPDU Sent             :0
          TCN: 0, Config: 0, RST: 0, MST: 0
 BPDU Received         :0
          TCN: 0, Config: 0, RST: 0, MST: 0
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page19 HUAWEI

STP端口信息显示：

此端口状态为Forwarding；

此端口角色为Root Port（根端口）；

端口默认优先级为128；

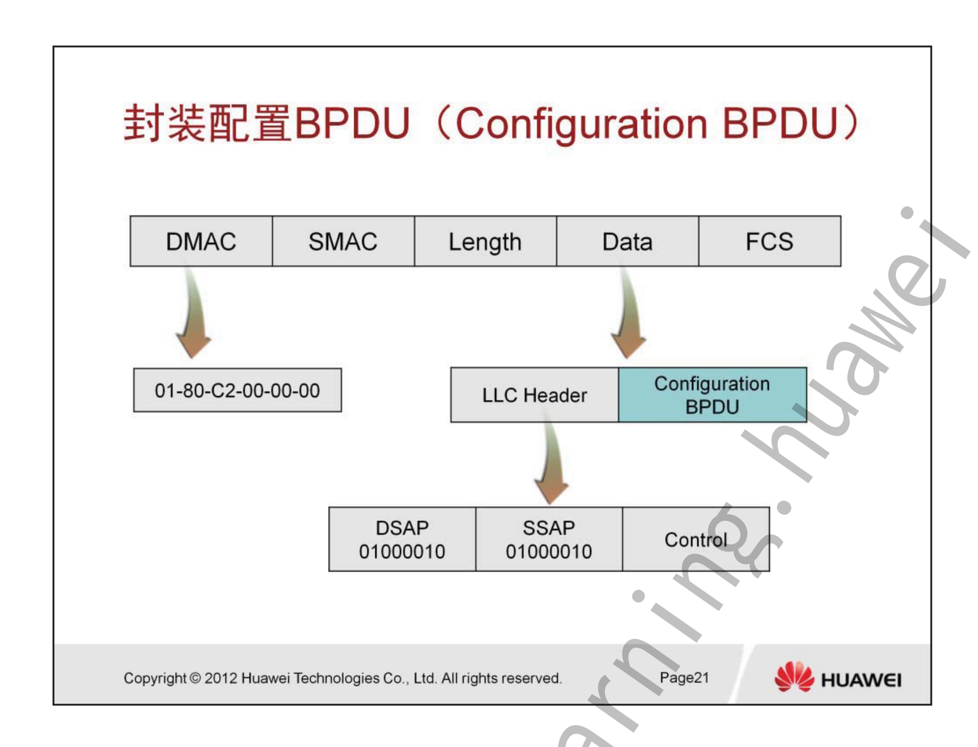
此端口所连网段的指定交换机为4096.00e0-fc41-4259，标识SWA。

更多资料获取：http://learning.huawei.com/cn



Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page20 HUAWEI

更多资料获取：http://learning.huawei.com/cn



用于计算生成树的各种信息和参数被封装在配置BPDU（Configuration Bridge Protocol Data Unit）中在交换机之间发送。

配置BPDU使用标准LLC格式封装在以太网数据帧中。

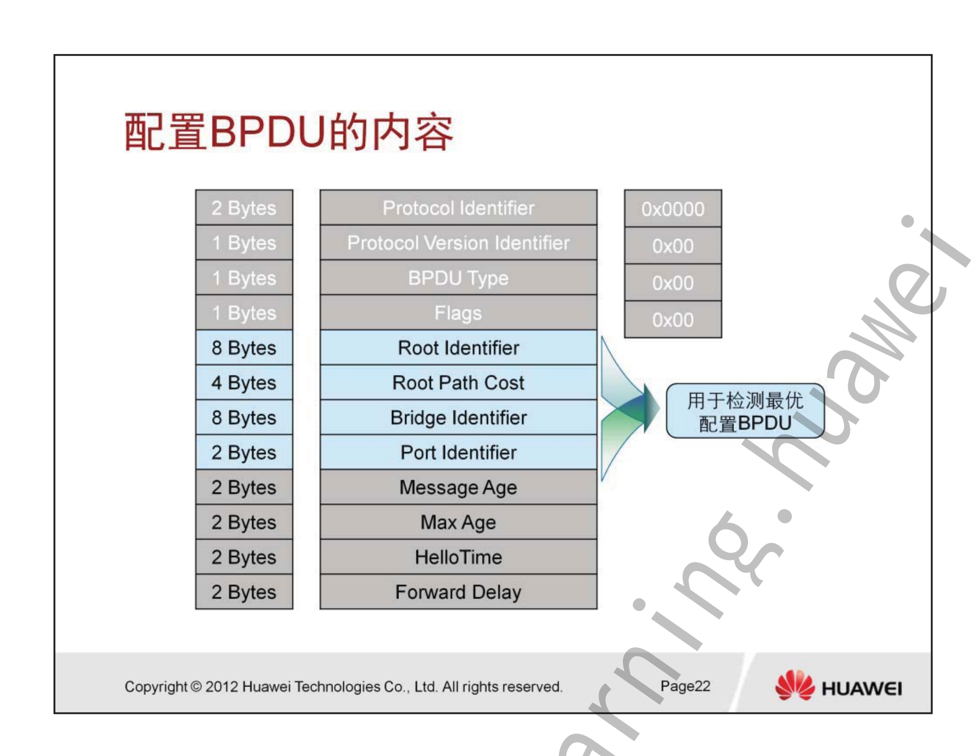
配置BPDU只在指定端口（Designated Port）上发送。

DMAC：目的MAC地址。发送配置BPDU的数据帧使用保留的组播MAC地址01-80-C2-00-00-00，此地址标识所有交换机，但是不能被交换机转发，也即只在本地链路有效。

LLC Header：目的服务访问点（Destination Service Access Point, DSAP）和源服务访问点（Source Service Access Point, SSAP）的值都设为二进制01000010。Control

字段的值设为3。

更多资料获取：http://learning.huawei.com/cn



当配置BPDU只用于计算生成树，不用于传递拓扑改变信息（第四章中详细描述）的时候：

Protocol Identifier（协议标识），Protocol Version Identifier（协议版本标识）和BPDU Type（BPDU类型）Flags（标志）四部分设置为全0。

Root Identifier，Root Path Cost，Bridge Identifier和Port Identifier四部分用于检测最优的配置BPDU，进行生成树计算。

Message Age随时间增长而变大；

Max Age默认为20秒，如果Message Age达到Max Age，则此配置BPDU被认为已经过期。

Hello Time默认为2秒，也即在指定端口上，配置BPDU每隔两秒发送一次。

Forward Delay默认为15秒。

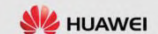
更多资料获取：http://learning.huawei.com/cn

配置BPDU中的重要参数

参 数	描 述
Root Identifier	发送此配置BPDU的交换机所认为的根交换机的交换机标识
Root Path Cost	从发送此配置BPDU的交换机到达根交换机的最短路径总开销，含交换机根端口的开销，不含发送此配置BPDU的端口的开销
Bridge Identifier	发送此配置BPDU的交换机的交换机标识
Port Identifier	发送此配置BPDU的交换机端口的端口标识

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page23



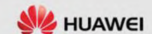
此表格列出了配置BPDU中四个与检测最优配置BPDU相关的参数以及相关描述。

更多资料获取：http://learning.huawei.com/cn

交换机全局参数

参数	描述
Designated Root	此交换机所认为的根交换机的交换机标识，用于设置此交换机所发送的配置BPDU中的Root Identifier参数
Root Path Cost	从此交换机到达根交换机的最短路径总开销，含此交换机的根端口的端口开销，用于设置此交换机所发送的配置BPDU中的Root Path Cost参数
Root Port	根端口的端口标识
Bridge Identifier	该交换机的交换机标识

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page24



此表格列出了重要的交换机全局参数，相关描述以及与配置BPDU中相关参数的关系。

根端口的选举与端口所维护的优先级参数有关，在后文中详细描述。

当交换机启动，初始化生成树协议时，Designated Root为交换机本身，也即交换机刚刚初始化的时候，总是认为自身为根交换机；Root Path Cost为0；没有Root Port，所有已被启用的端口都是指定端口（Designated Port）。

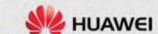
因此，当交换机初始化之后，从所有端口上向外发送Root Identifier为自身标识，Root Path Cost为0，Bridge Identifier为自身标识，Port Identifier为发送端口的端口标识的配置BPDU。

更多资料获取：http://learning.huawei.com/cn

端口参数

参　数	描　述
Path Cost	本地端口开销，只对根端口有意义
Designated Root	从该端口上所接收到的配置BPDU中所记录的根交换机的交换机标识
Designated Cost	指定开销，对于指定端口，该参数等于该交换机的Root Path Cost；对于非指定端口，该参数等于上行交换机发送的配置BPDU中设置的Root Path Cost
Designated Bridge	指定交换机，对于指定端口，该参数标识该端口所属的交换机；对于非指定端口，该参数标识该端口所连网段上的指定交换机
Designated Port	指定端口，对于指定端口，该参数标识自身端口；对于非指定端口，该参数标识该端口所连网段上的指定端口

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page25



此表列出了端口上维护的重要参数，这些参数大都和配置BPDU有关系。

当交换机初始化之后，由于认为自身为根交换机，所有端口均为指定端口，因此端口上的参数设置如下：

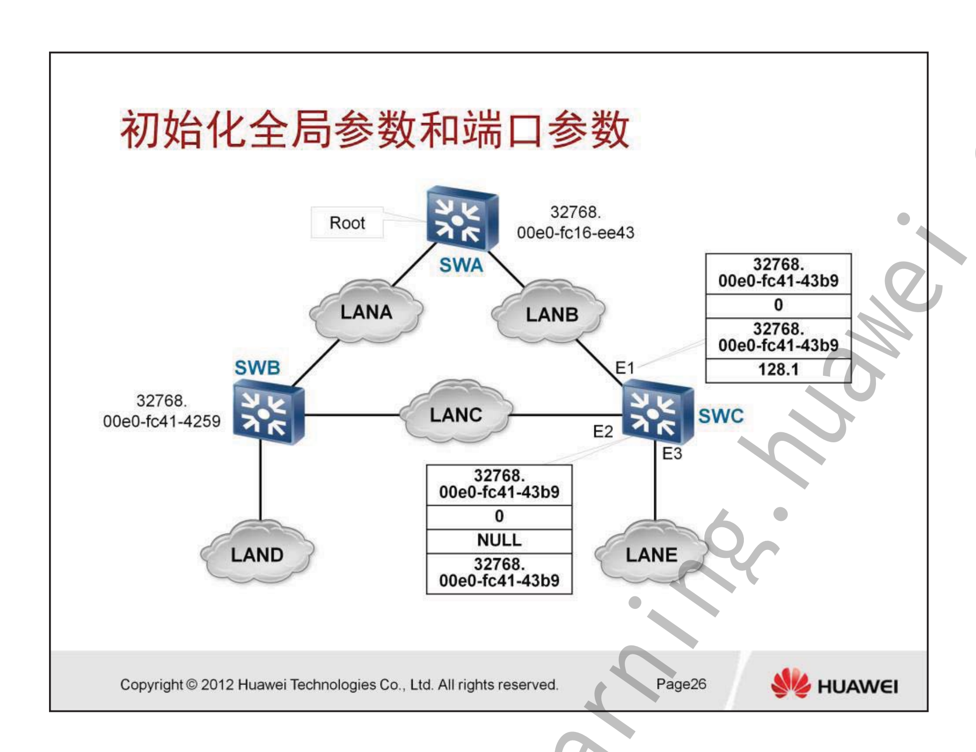
Designated Root为交换机本身；

Designated Cost为0；

Designated Bridge为交换机本身；

Designated Port为端口自身的标识。

更多资料获取：http://learning.huawei.com/cn



本例中：

SWA为网络中的原有根交换机（Root），SWC为刚加入到该网络中的新交换机。

SWC初始化全局参数为：

Designated Root为自身的交换机标识32768.00e0-fc41-43b9；

Root Path Cost为0；

Root Port为空；

Bridge Identifier为自身的交换机标识32768.00e0-fc41-43b9。

端口参数初始化为：

Designated Root和Designated Bridge初始化为交换机本身32768.00e0-fc41-43b9；

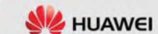
Designated Cost初始化为0；

Designated Port初始化为自身的端口标识－默认优先级128与端口号的组合。

更多资料获取：http://learning.huawei.com/cn

收到更优配置BPDU并记录在端口参数中

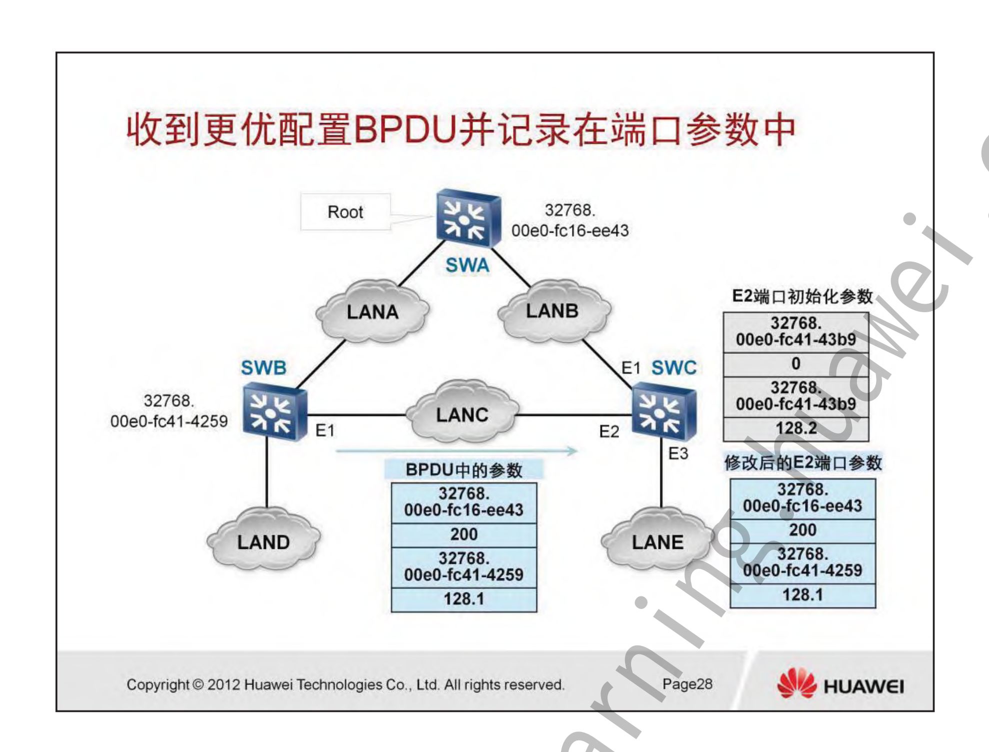
比较次序	比较内容
1	比较所接收到的配置BPDU中Root Identifier和端口参数中记录的Designated Root，如果二者相等则进入第二步
2	比较所接收到的配置BPDU中Root Path Cost和端口参数中记录的Designated Cost，如果二者相等则进入第三步
3	比较所接收到的配置BPDU中Bridge Identifier和端口参数中记录的Designated Bridge，如果二者相等则进入第四步
4	比较所接收到的配置BPDU中Port Identifier和端口参数中记录的Designated Port

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page27 

端口上的参数如前所述进行初始化之后，开始比较在端口上所接收到的配置BPDU和端口参数中的记录，如果端口参数中记录的更优先（表中所列出的比较项均为值越小越优先），则丢弃配置BPDU，如果配置BPDU中记录的更优先，则对端口参数做如下修改：

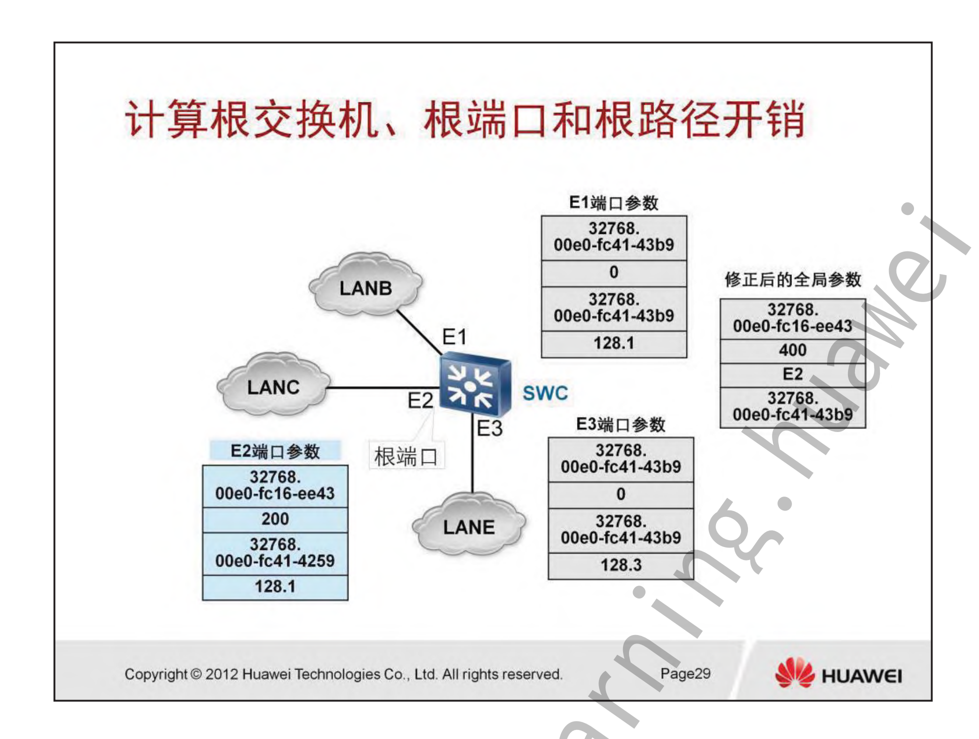
端口参数中的Designated Root、Designated Cost、Designated Bridge和Designated Port分别设置成此配置BPDU中的Root Identifier、Root Path Cost、Bridge Identifier和Port Identifier。

更多资料获取：http://learning.huawei.com/cn



假设初始化之后SWC的E2端口首先从SWB收到一个配置BPDU，SWC收到此配置BPDU后，将此配置BPDU中的相关参数和E2端口的参数相比较，发现配置BPDU中的Root Identifier比E2端口参数中记录的Designated Root更优先，按照如前所述的规则，将E2端口的相关参数修改为配置BPDU中的值。

更多资料获取：http://learning.huawei.com/cn



从端口上收到一个更优的配置BPDU之后，重新计算根交换机、根端口和根路径开销的过程将被启动。

计算过程为：

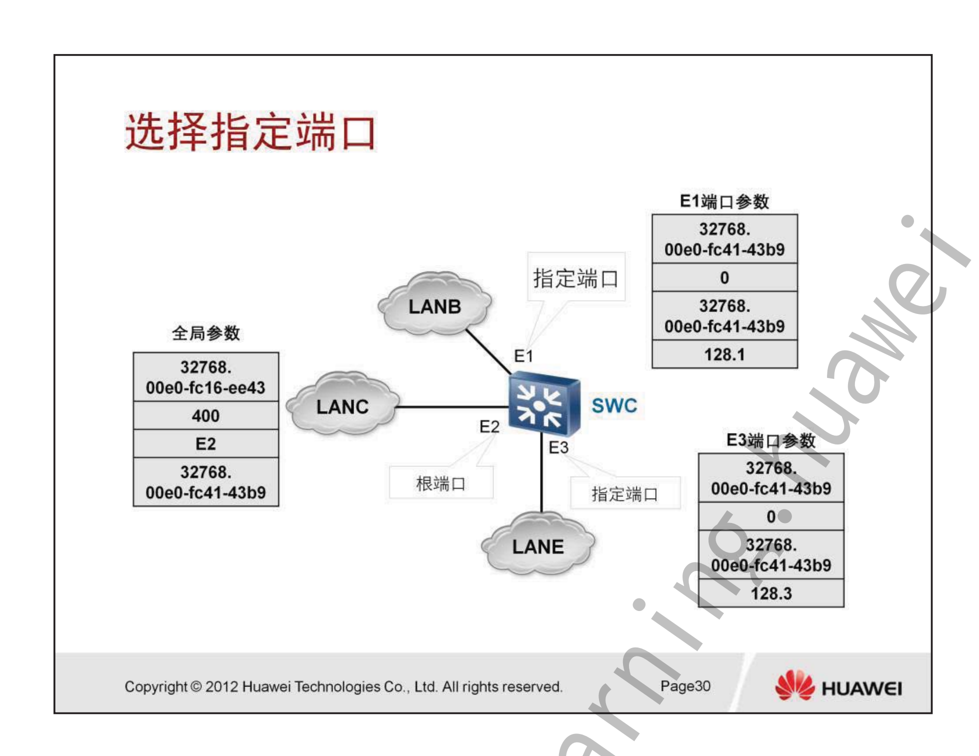
1. 根据所有的端口上记录的参数，依次比较Designated Root，Designated Cost和端口Cost之和，Designated Bridge和Designated Port，从中选出一个记录了最优参数的端口，并且此端口上记录的Designated Root要比交换机自身的Bridge Identifier（交换机标识）更优先，此端口即为根端口；

2. 选择出此端口之后，更新交换机全局参数Designated Root为根端口记录的Designated Root；更新交换机全局参数Root Path Cost为根端口记录的Designated Cost与根端口的Port Cost之和；

3. 如果任何端口记录的Designated Root参数都不比交换机自身的Bridge Identifier更优先，则交换机全局参数Designated Root设置为交换机自身的Bridge Identifier；交换机全局参数Root Path Cost设置为0。

本例中：E2端口记录了最优的参数，因此更新全局参数如图所示，Designated Root为E2端口记录的Designated Root；Root Path Cost为E2端口记录的Designated Cost（200）和E2端口的Port Cost（200）之和（400）。

更多资料获取：http://learning.huawei.com/cn



根交换机，根端口，根路径开销计算完成，并更新了交换机全局参数之后，启动指定端口选择过程，决定非根端口是否可以成为指定端口。

满足下列条件之一的，成为所连网段上的指定端口：

1. 已经被选为指定端口，即端口所记录的Designated Bridge为交换机自身的交换机标识，Designated Port为端口自身的端口标识；

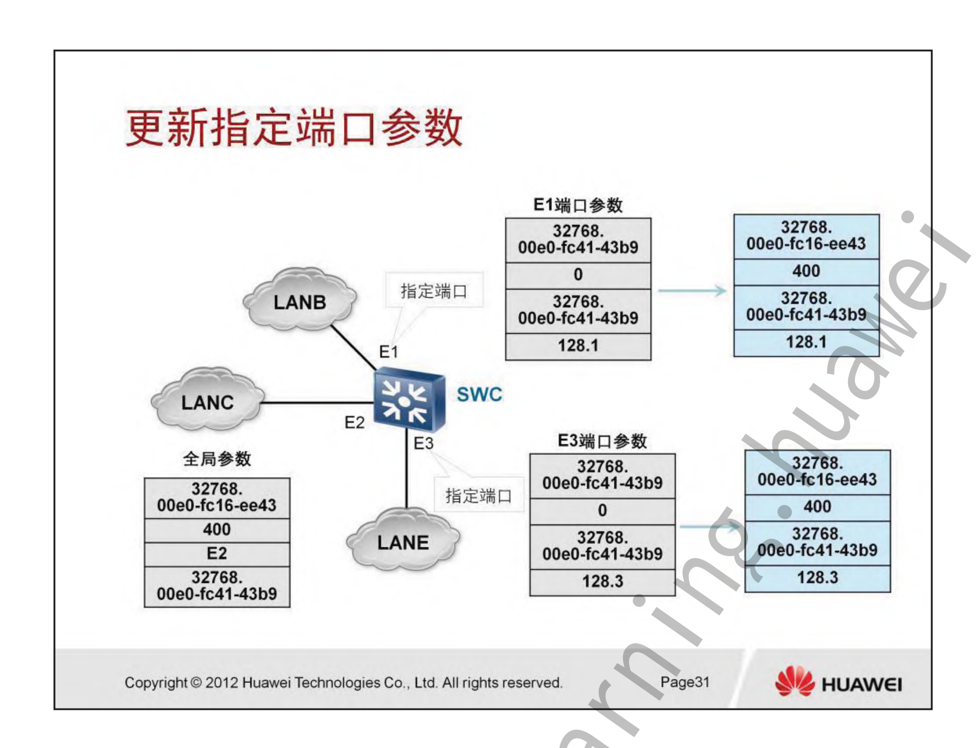
2. 端口参数中的Designated Root和交换机全局参数Designated Root不一致；

3. 交换机全局参数Designated Root和端口参数Designated Root一致，但是交换机全局参数Root Path Cost比端口参数Designated Cost更优先；

4. 全局参数中的根交换机和根路径开销都和端口记录的一致，但是交换机自身的交换机标识比端口记录的Designated Bridge更优先；

5. 全局参数中的根交换机和根路径开销都和端口记录的一致，交换机自身的交换机标识和端口记录的Designated Bridge一致，但是端口自身的标识比端口记录的Designated Port更优先。

更多资料获取：http://learning.huawei.com/cn



确定一个端口可以成为指定端口之后，交换机需要修改指定端口的参数，修改规则如下：

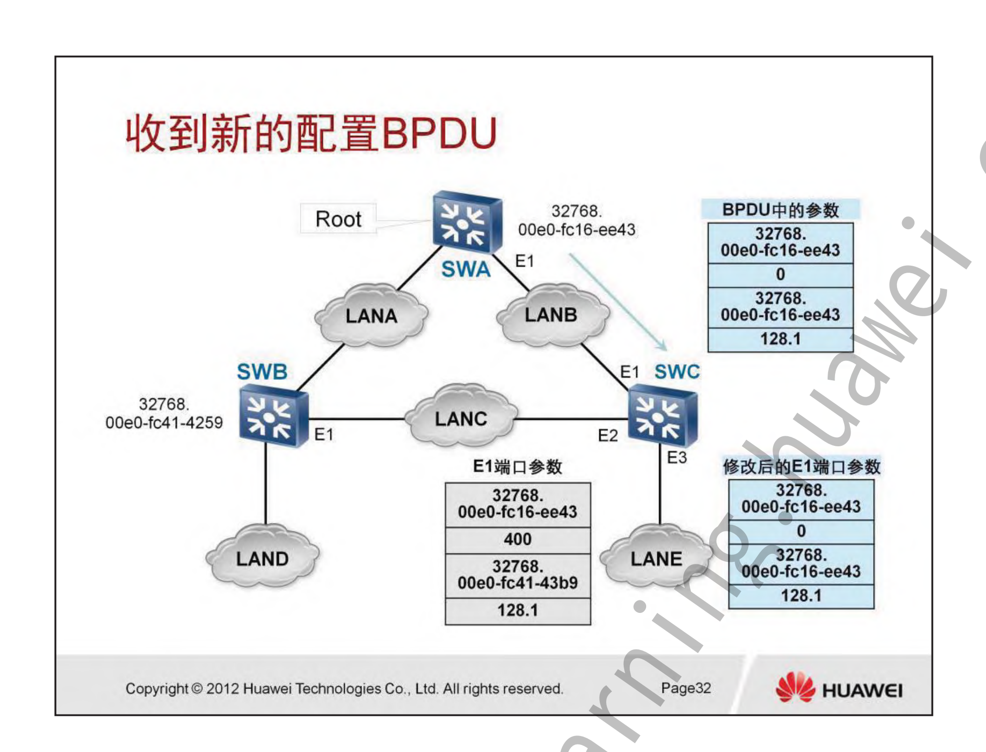
Designated Root设置为交换机全局参数Designated Root；

Designated Cost设置为交换机全局参数Root Path Cost；

Designated Bridge设置为交换机自身的Bridge Identifier；

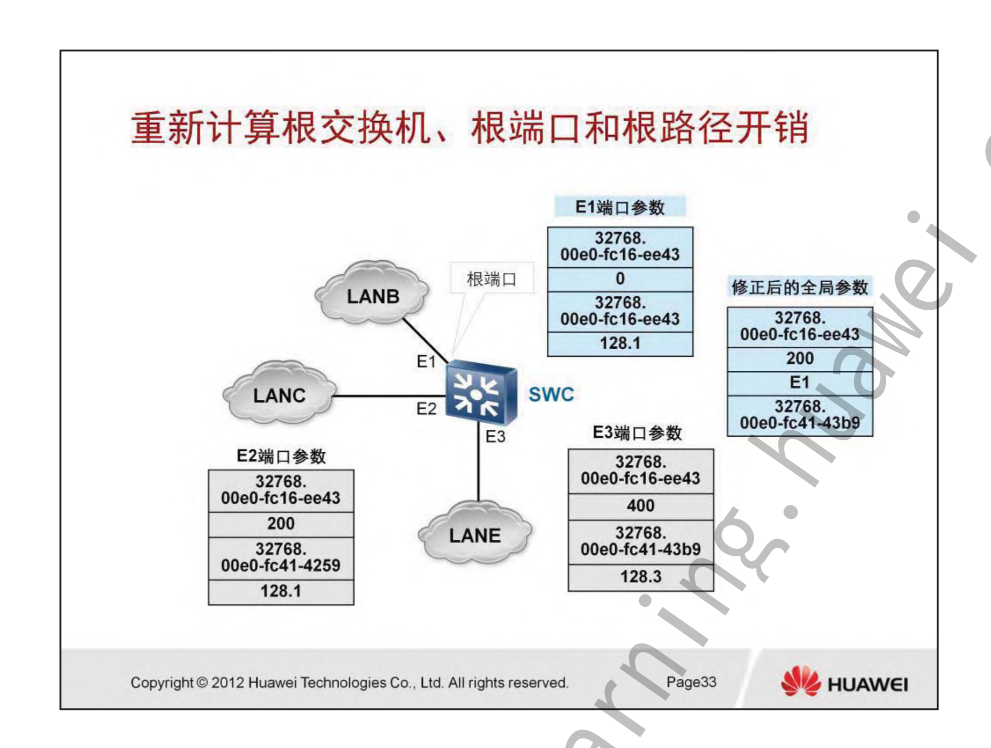
Designated Port设置为端口自身的Port Identifier。

更多资料获取：http://learning.huawei.com/cn



假设处理完E2端口上收到的配置BPDU之后，SWC又从E1端口上收到SWA发送的配置BPDU，处理过程如前所述，首先比较端口参数和配置BPDU中的参数，发现新收到的BPDU比端口记录的参数更优先，因此修改E1端口的参数。

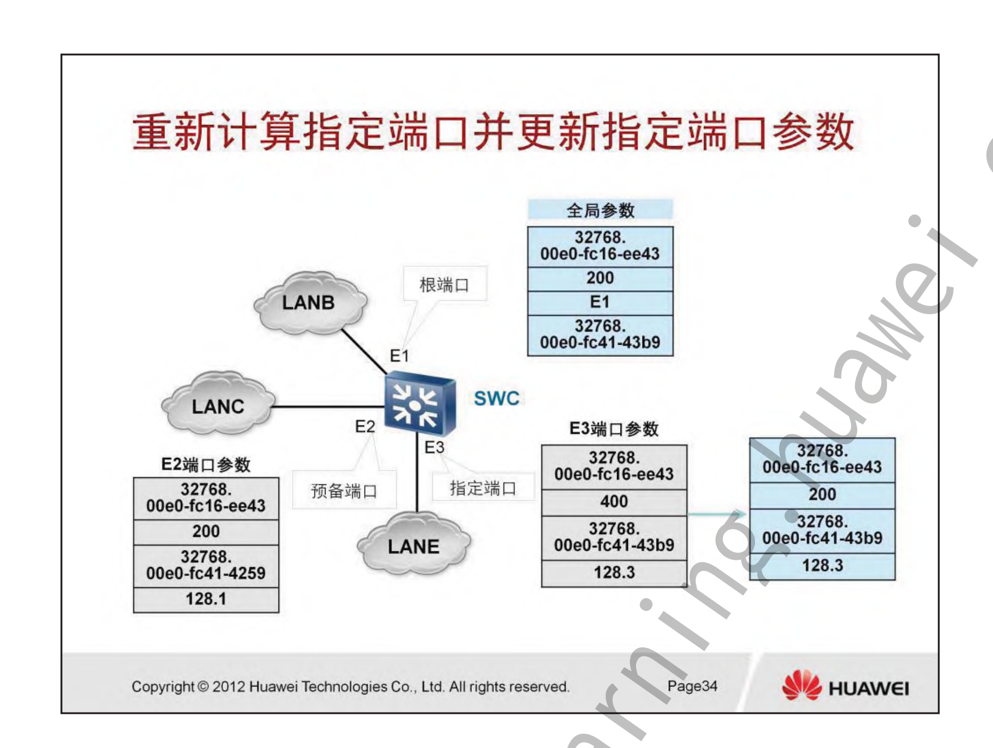
更多资料获取：http://learning.huawei.com/cn



根据如前所述的规则，重新启动计算根交换机、根端口和根路径开销的过程，E1端口被选择为新的根端口。

选择新的根端口之后，重新修改交换机全局参数，Root Path Cost修改为200，根端口更新为E1。

更多资料获取：http://learning.huawei.com/cn



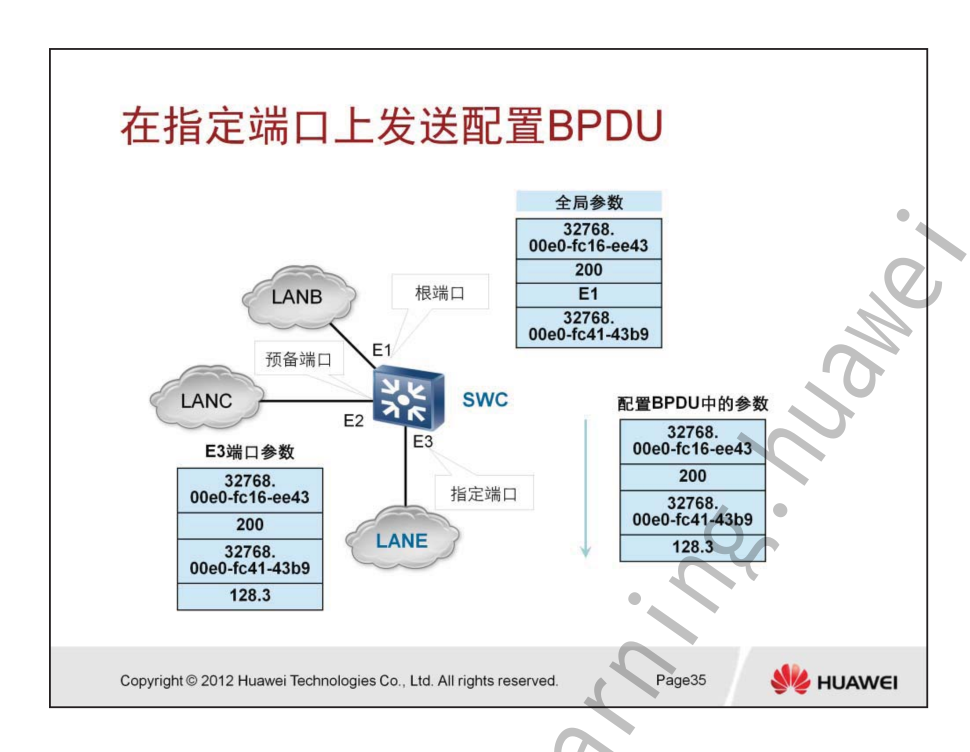
按照如前所述规则，检测除了新的根端口之外的其它端口是否可以成为指定端口，不能成为指定端口的，成为预备端口（Alternate Port）。

本例中：

E3端口符合成为指定端口的条件（如前所述条件列表第一条），因此E3端口成为指定端口；

E2端口不符合成为指定端口的任何条件，因此E2端口成为预备端口（Alternate Port）。

更多资料获取：http://learning.huawei.com/cn



交换机在指定端口上向外发送配置BPDU。

配置BPDU中的参数设置规则如下：

1. Root Identifier设置为端口参数Designated Root，也即全局参数Designated Root；

2. Root Path Cost设置为端口参数Designated Cost，也即全局参数Root Path Cost；

3. Bridge Identifier设置为交换机自身的Bridge Identifier；

4. Port Identifier设置为发送端口的Port Identifier。

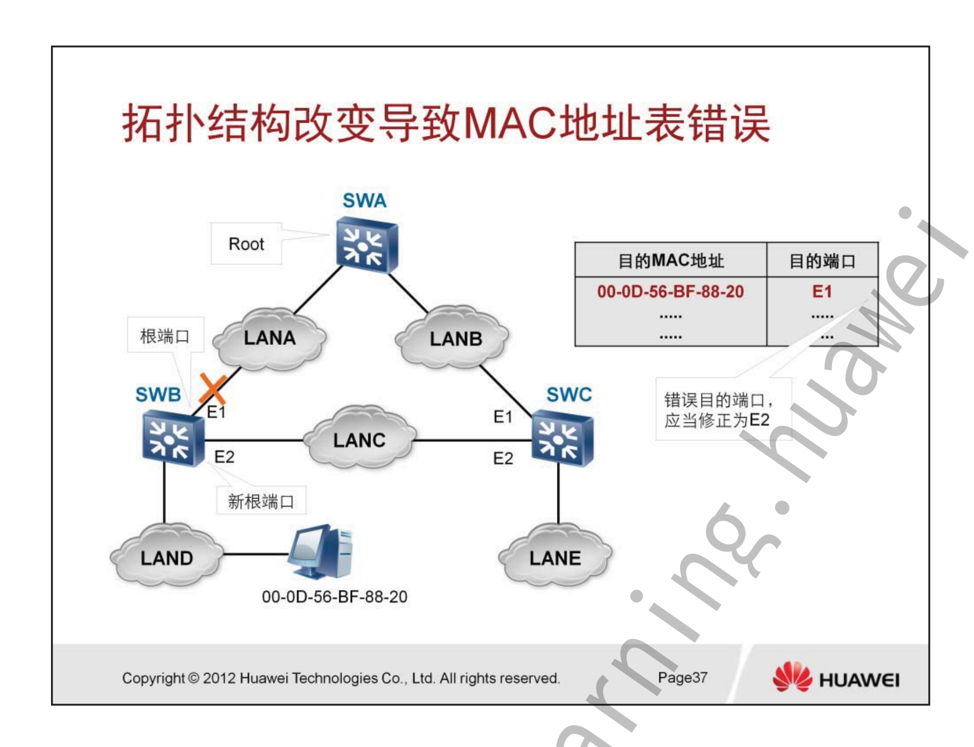
更多资料获取：http://learning.huawei.com/cn



Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page36 HUAWEI



更多资料获取：http://learning.huawei.com/cn



默认情况下，MAC地址表中的动态表项生存期为300秒（5分钟）。

本例中：

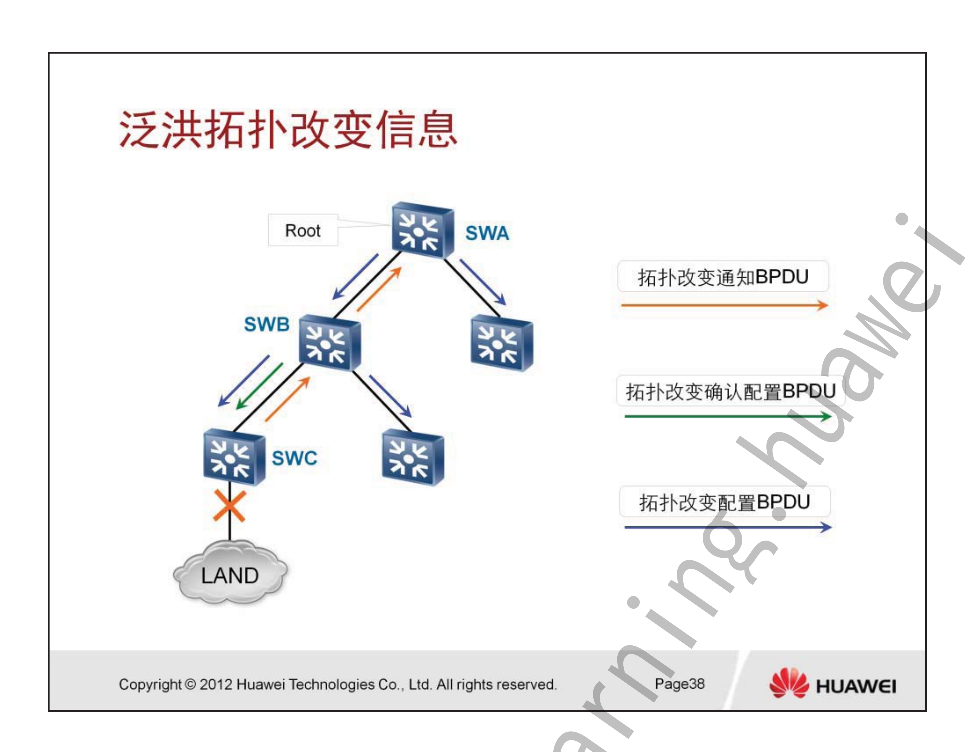
稳定拓扑下，在SWC上到达LAND某PC的目的端口应当为E1；

当SWB的E1接口断开之后，E2接口成为新的根端口，从SWC到达该PC的目的地址应当修改为E2，但是交换机不能检测到拓扑改变，导致MAC地址表错误，最长可导致5分钟的数据转发错误。

解决问题的办法：当拓扑结构改变之后，通过一定的机制，使拓扑改变的信息在整网内泛洪，并修改MAC地址表的生存期为一个较短的数值，等拓扑结构稳定之后，再恢复MAC地址表的生存期。

STP规定这个较短的MAC地址表生存期使用交换机的Forward Delay参数，默认为15秒。

更多资料获取 http://learning.huawei.com/cn



在向整网泛洪拓扑改变信息的过程中，共涉及三种BPDU：

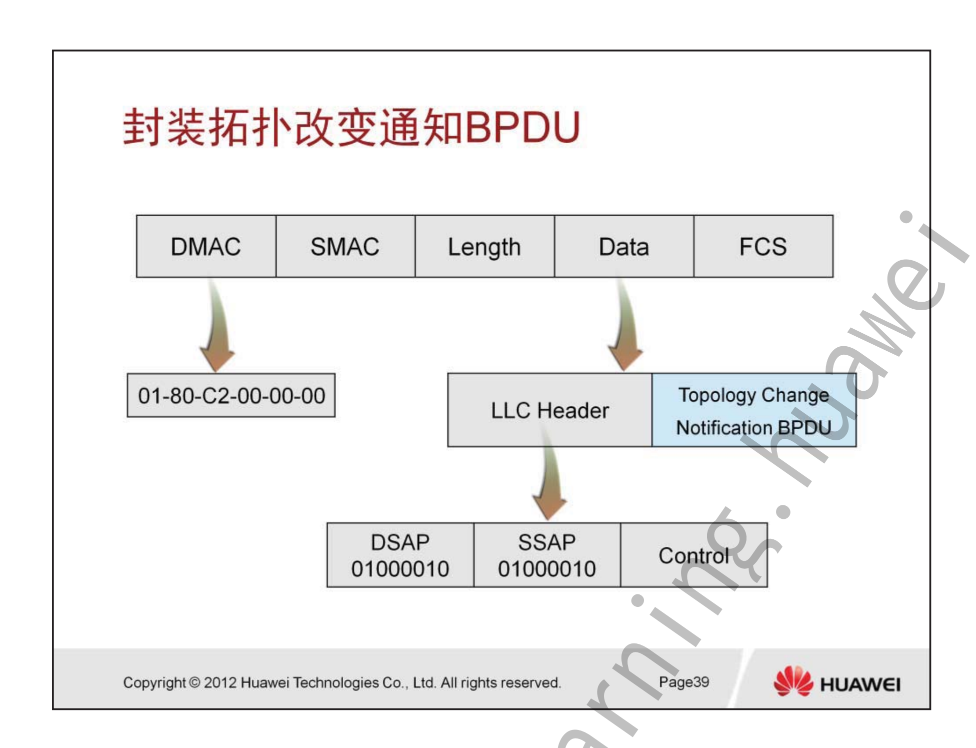
1. 拓扑改变通知BPDU：Topology Change Notification BPDU。用于非根交换机在根端口上向上行交换机通告拓扑改变信息，并且每隔Hello Time（2秒）发送一次，直到收到上行交换机的拓扑改变确认配置BPDU或者拓扑改变配置BPDU。

2. 拓扑改变确认配置BPDU：Topology Change Acknowledgment Configuration BPDU。配置BPDU的一种，和普通配置BPDU不同的是此配置BPDU设置了一个Flag位。用于非根交换机在接收到拓扑改变通知BPDU的指定接口上向下行交换机发送拓扑改变通知的确认信息。

3. 拓扑改变配置BPDU：Topology Change Configuration BPDU。此配置BPDU设置了另外一个Flag位。用于从根交换机向整网泛洪拓扑改变信息，所有交换机都在自己所有的指定端口上泛洪此BPDU。

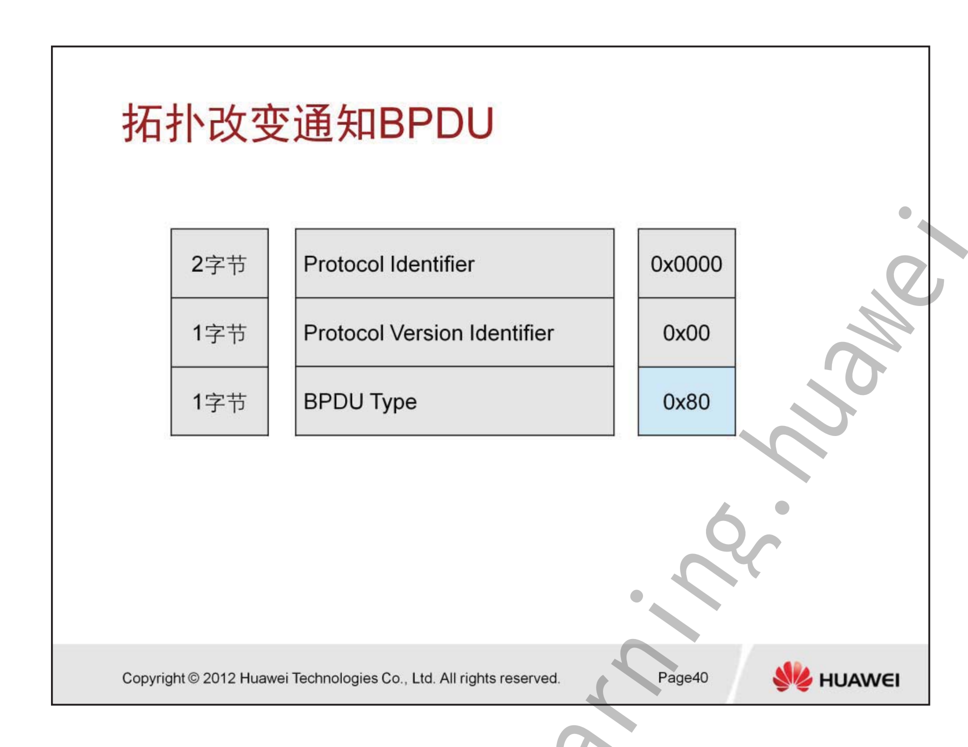
SWA收到SWB发送的拓扑改变通知BPDU之后，每隔2秒向网络中发送拓扑改变配置BPDU（设置了一个Flag位的配置BPDU），使网络中所有的交换机都把MAC地址表的生存期修改为Forward Delay（15秒），经过一段时间（Max Age加上Forward Delay，默认为35秒）之后，SWA（根交换机）在自己发送的配置BPDU中，清除Flag位，表示网络拓扑已经稳定，网络中的交换机恢复MAC地址生存期。

更多资料获取：http://learning.huawei.com/cn



拓扑改变通知BPDU和配置BPDU使用的是相同的封装方式。

更多资料获取：http://learning.huawei.com/cn



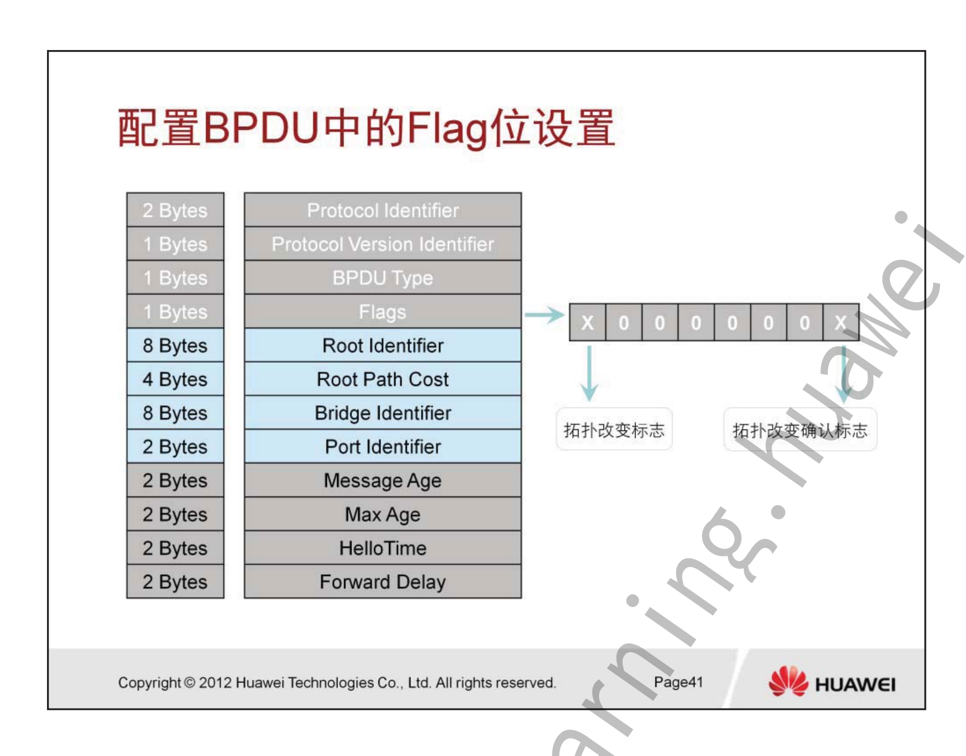
拓扑改变通知BPDU的格式比较简单，没有配置BPDU中那么多参数。

Protocol Identifier设置为全0；

Protocol Version Identifier设置为全0；

BPDU Type设置为二进制1000 0000，即十六进制0x80。

更多资料获取：http://learning.huawei.com/cn



拓扑改变确认配置BPDU和拓扑改变配置BPDU都是配置BPDU的一种，和普通的配置BPDU不同的是：

普通的配置BPDU中Flag字段全部设置为0；

拓扑改变确认配置BPDU将Flag字段的第8位设置为1；

拓扑改变配置BPDU将Flag字段的第1位设置为1。

更多资料获取：http://learning.huawei.com/cn



问 题

生成树协议如何在网络中计算出一棵无环的树?

生成树协议如何解决临时环路问题?

生成树协议如何解决拓扑改变引起的MAC地址表错误问题?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page42  

生成树协议如何在网络中计算出一棵无环的树?

在网络中选出一个根交换机，为每个非根交换机选择一个根端口，为每个网段选择一个指定端口，将既不是根端口也不是指定端口的端口设为阻塞状态。

生成树协议如何解决临时环路问题?

当一个端口从不转发状态转为转发状态之前，要经过两个Forward Delay间隔，以确保网络中其它交换机完成生成树计算。

生成树协议如何解决拓扑改变引起的MAC地址表错误问题?

拓扑改变之后，拓扑改变信息在整网内泛洪，交换机将MAC地址表生存期设置为一个较短的数值，拓扑结构稳定之后，交换机再恢复MAC地址表的生存期。

http://learning.huawei.com/cn 更多资料获取

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

本课程介绍RSTP（快速生成树协议）的原理和配置。

RSTP是STP的升级版本，与STP相比，最显著的特点就是通过新的机制，加快了收敛速度。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 描述RSTP的基本计算过程
- 描述RSTP端口状态的迁移
- 描述拓扑结构改变信息的泛洪过程

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page2 HUAWEI

更多资料获取：http://learning.huawei.com/cn



目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

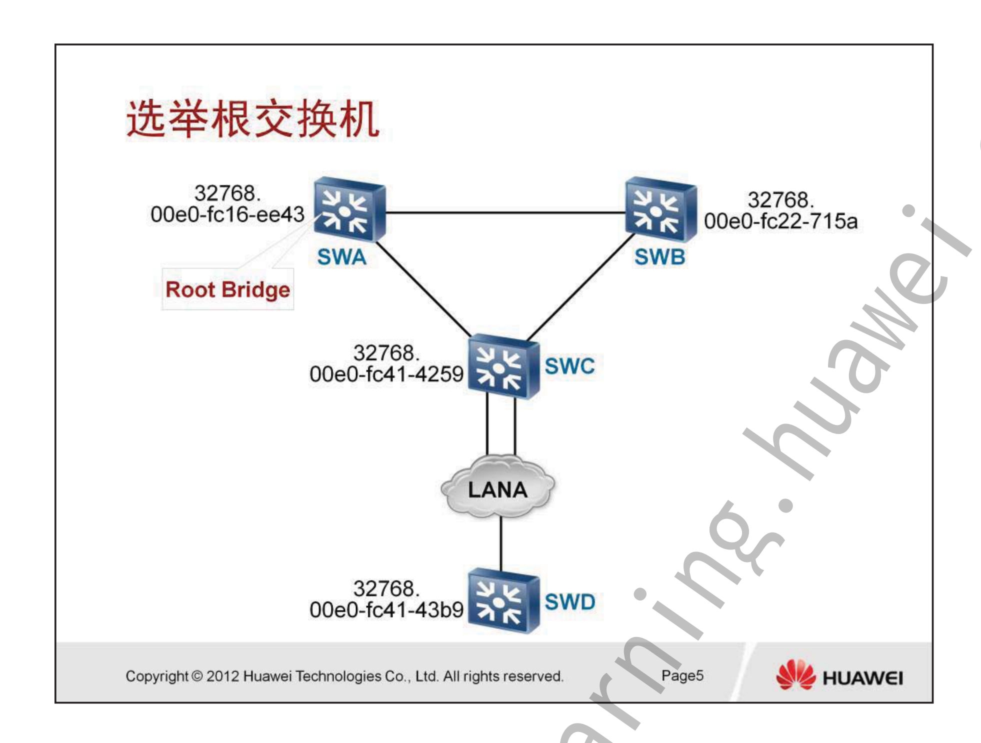
Page3

HUAWEI

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



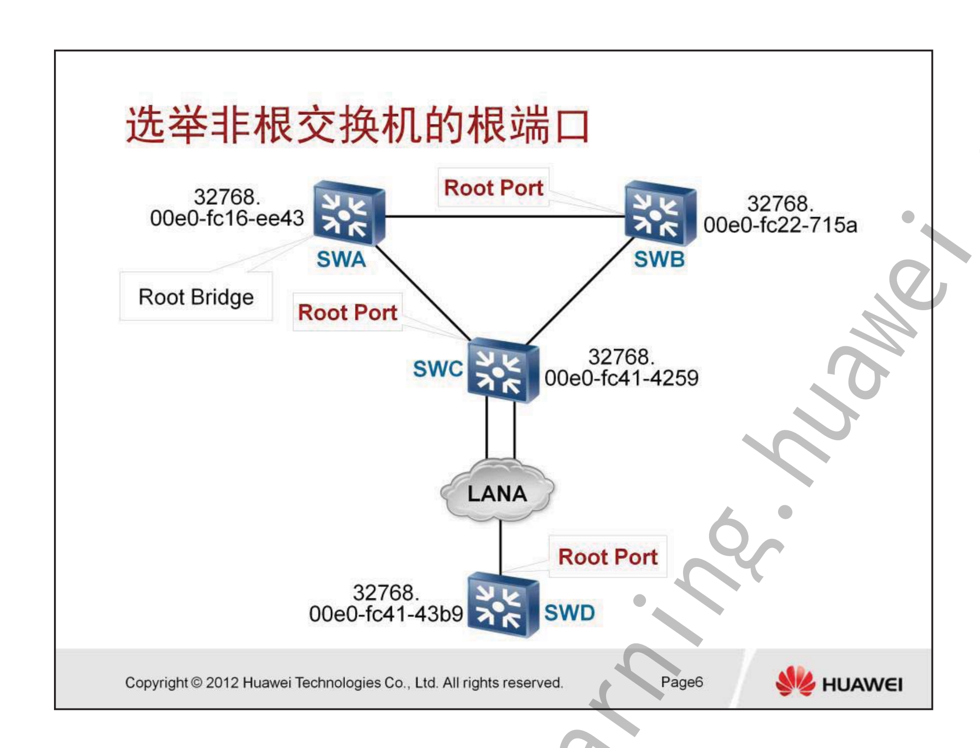
如同STP的基本计算过程，RSTP计算的第一步也是选举根交换机（Root Bridge），根交换机的选举基于交换机标识（Bridge Identifier）。

交换机标识由两部分组成：两字节长度的交换机优先级和六字节长度的MAC地址。

交换机优先级是可以配置的，取值范围是0～65535，默认值为32768。网络中交换机标识最小的成为根交换机，首先比较优先级，如果优先级相同则比较MAC地址，值越小越优先。

本例中，三个交换机的优先级是相同的，由于SWA的MAC地址值最小，因此SWA为根交换机。

更多资料获取：http://learning.huawei.com/cn



RSTP为每个非根交换机选举根端口（Root Port）。

交换机的每个端口都有一个端口开销（Port Cost）的参数，此参数表示数据从该端口发送时的开销值，也即出端口的开销。RSTP认为从一个端口接收数据是没有开销的。

端口的开销和端口的带宽有关，带宽越高，开销越小，VRP平台中，百兆端口的802.1T开销值为199999。

从一个非根交换机到达根交换机的路径可能有多条，每一条路径都有一个总的开销值，此开销值是该路径上所有出端口的端口开销总和。

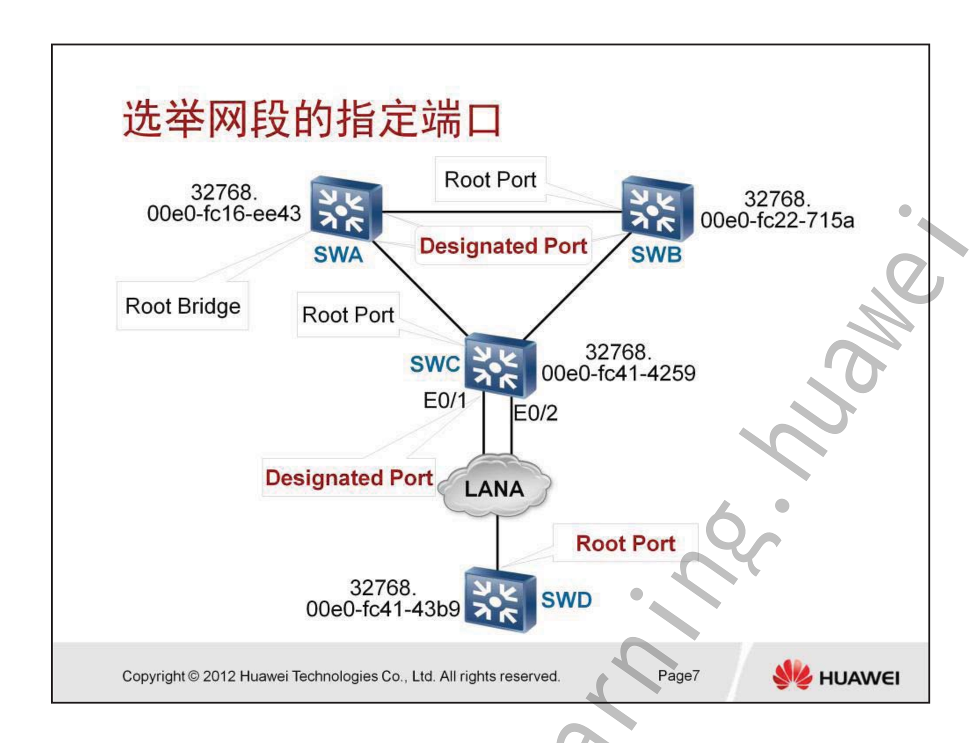
根端口是指从一个非根交换机到根交换机总开销最小的路径所经过的本地端口。这个最小的总开销值称为交换机的根路径开销（Root Path Cost）。如果这样的端口有多个，则比较端口上所连接的上行交换机的交换机标识，越小越优先，如果端口上所连接的上行交换机的交换机标识相同，则比较端口上所连接的上行端口的端口标识（Port Identifier），越小越优先。

端口标识由两部分组成：一字节长度的端口优先级和一字节长度的端口号。

一字节长度的端口优先级是可配置的，默认为128。

本例中，假设所有端口都是百兆端口，使用相同的开销值199999。

更多资料获取：http://learning.huawei.com/cn

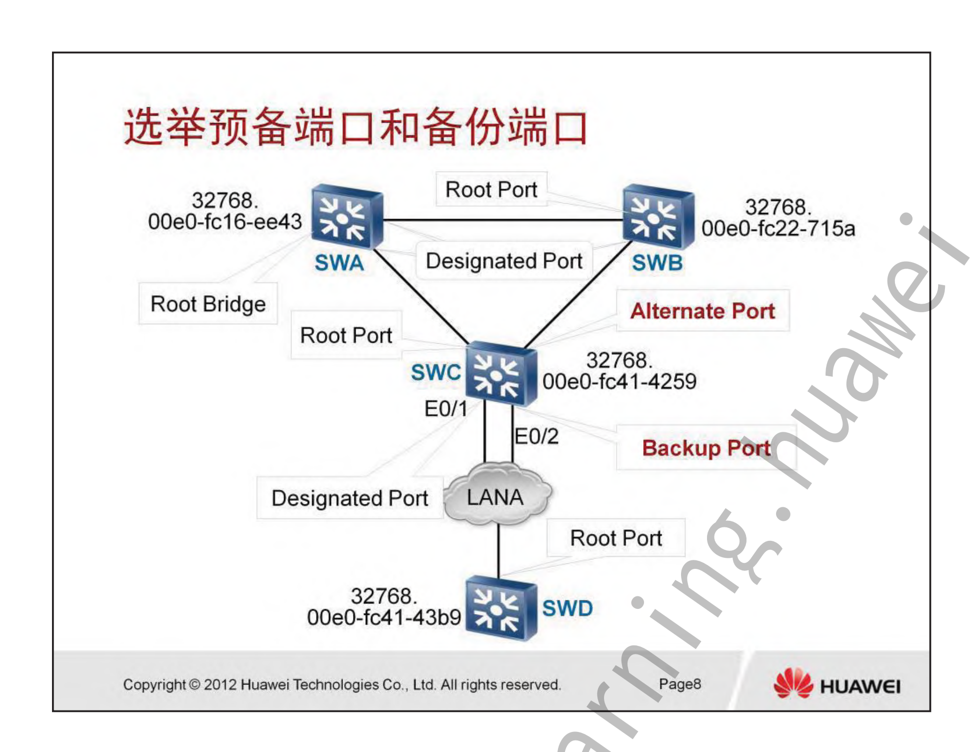


RSTP为每个网段选出一个指定端口（Designated Port），指定端口为每个网段转发发往根交换机方向的数据，并且转发由根交换机方向发往该网段的数据。指定端口所在的交换机称为该网段的指定交换机。

为每个网段选举指定端口和指定交换机的时候，首先比较该网段所连接的端口所属交换机的根路径开销，越小越优先；如果根路径开销相同，则比较所连接的端口所属交换机的交换机标识，越小越优先；如果根路径开销相同，交换机标识也相同，则比较所连接的端口的端口标识，越小越优先。

本例中，SWA为根交换机，因此所有端口均为指定端口；对于SWC和SWB之间的链路，由于该链路所连接的两个交换机SWC和SWB的根路径开销相同，因此比较两个交换机的标识，SWB的交换机标识较小，因此指定端口在SWB上；对于LANA，由于SWC的根路径开销小于SWD的根路径开销，所以指定交换机为SWC，由于SWC有两个端口连接到LANA，因此，比较两个端口的端口标识，默认端口优先级相同，为128，由于E0/1的端口号较小，因此LANA的指定端口是SWC的E0/1接口。

更多资料获取：http://learning.huawei.com/cn



对于既不是根端口，也不是指定端口的交换机端口：

如果该端口属于所连接网段的指定交换机，则端口状态设置为备份端口（Backup Port）；

如果该端口不属于所连接网段的指定交换机，则端口状态设置为预备端口（Alternate Port）。

预备端口主要是为了备份根端口，而备份端口主要是为了备份指定端口。

无论是备份端口还是预备端口，都不处于转发状态。

更多资料获取：http://learning.huawei.com/cn

交换机端口角色

端口角色	描述
Root Port	根端口，是所在交换机上离根交换机最近的端口，稳定时处于转发状态。
Designated Port	指定端口，转发所连接的网段发往根交换机方向的数据和从交换机方向发往所连接的网段的数据，稳定时处于转发状态。
Backup Port	备份端口，不处于转发状态，所属交换机为端口所连网段的指定交换机。
Alternate Port	预备端口，不处于转发状态，所属交换机不是端口所连网段的指定交换机。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page9



如前所述，对于物理层和数据链路层可以正常工作，并且开启了RSTP的交换机端口，RSTP共定义了四种端口角色，稳定时处于转发状态的有根端口和指定端口。

底层没有开启的端口称为Disable端口。

更多资料获取：http://learning.huawei.com/cn



Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

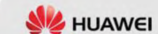
HUAWEI

更多资料获取：http://learning.huawei.com/cn

交换机端口状态

端口状态	描述
Discarding 丢弃状态	此状态下端口对接收到的数据做丢弃处理，端口不转发数据帧，不学习MAC地址表。 Alternate Port和Backup Port
Learning 学习状态	此状态下端口不转发数据帧，但是学习MAC地址表，参与计算生成树，接收并发送BPDU。
Forwarding 转发状态	此状态下端口正常转发数据帧，学习MAC地址表，参与计算生成树，接收并发送BPDU。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page11



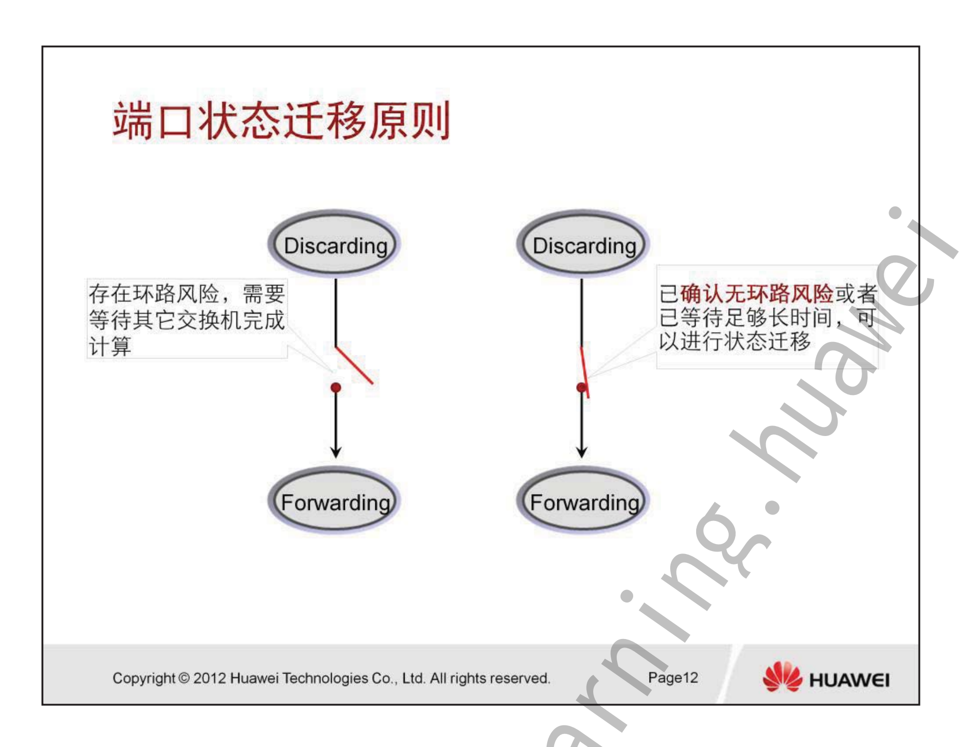
与STP不同，RSTP只定义了三种端口状态：Discarding（丢弃）状态，Learning（学习）状态，Forwarding（转发）状态。

预备端口（Alternate Port）和备份端口（Backup Port）处于Discarding状态；

指定端口（Designated Port）和根端口（Root Port）稳定情况下处于Forwarding状态；

Learning状态是指定端口和根端口在进入转发状态之前的一种临时状态。

更多资料获取：http://learning.huawei.com/cn



根据选举规则，确定端口角色之后，需要根据端口角色设置端口状态。

将端口状态从Forwarding状态迁移到Discarding状态（从根端口或者指定端口变成预备端口或者备份端口）是不会出现环路风险的，可以不经过等待立即转换；

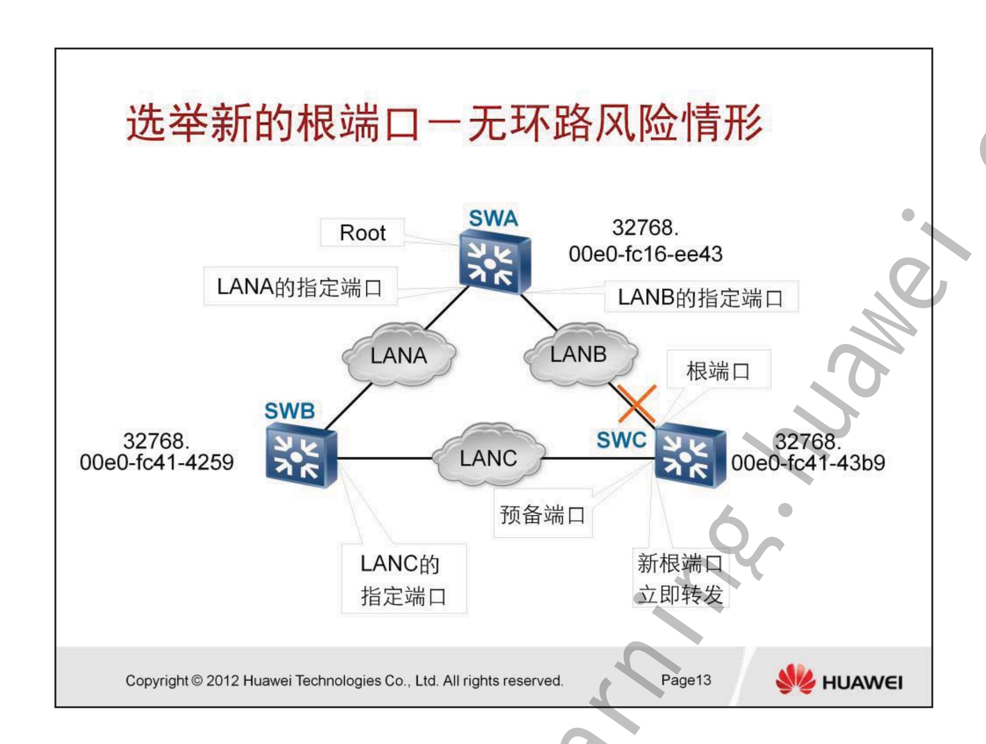
将端口状态从Forwarding状态迁移到Forwarding状态（从根端口变成指定端口或者从指定端口变成根端口）也不会引起环路风险，也可以不经过等待立即转换；

端口状态迁移时能引起环路风险的是从Discarding状态迁移到Forwarding状态（从预备端口或者备份端口变成根端口或者指定端口），在STP中，从不转发状态迁移到Forwarding状态需要等待两次Forward Delay间隔才能迁移，以保证网络中需要进入不转发状态的端口有足够的时间完成计算。但是RSTP对此做了改进。

RSTP（快速生成树）的主要设计原则是，在没有临时环路风险的情况下，使原本处于不转发状态下的端口在成为指定端口或根端口之后，尽可能快的进入Forwarding状态，加快收敛速度。

因此，如何确认网络中有没有环路风险是RSTP的重要内容。

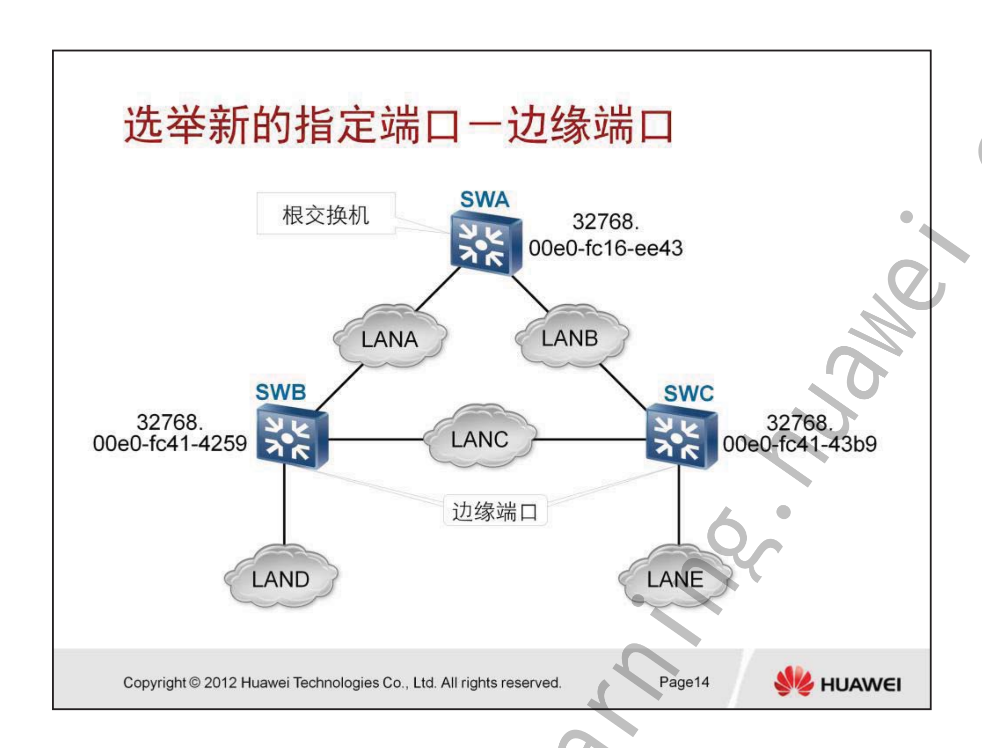
更多资料获取：http://learning.huawei.com/cn



一个非根交换机选举出一个新的根端口之后，如果以前的根端口已经不处于Forwarding状态，则新的根端口立即进入转发状态。

本例中：SWC上与LANB相连的端口为根端口，假设此端口断开，即不再处于转发状态，则SWC需要重新选择一个根端口，与LANC相连的端口于是从预备端口成为新的根端口。由于旧的根端口已经不再处于转发状态，因此网络中没有环路风险，新的根端口可以立即进入转发状态。

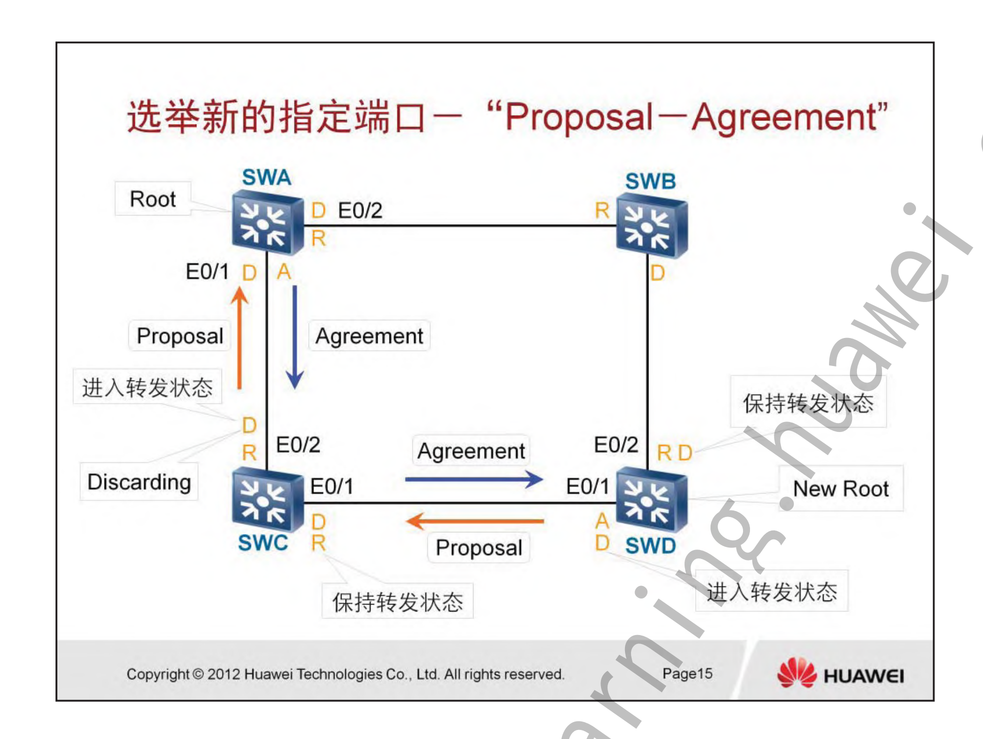
更多资料获取：http://learning.huawei.com/cn



边缘端口（Edge Port）是指不连接任何交换机的端口。

当把一个交换机端口配置成为边缘端口之后，一旦端口被启用，则端口立即成为指定端口（Designated Port），并进入转发状态。

更多资料获取：http://learning.huawei.com/cn



RSTP使用 “Proposal – Agreement”协商机制加快非边缘端口成为新的指定端口之后，从Discarding状态进入Forwarding状态的速度。

本例中，假设最初网络中各交换机的优先级优先次序为SWA>SWB>SWC>SWD；因此SWA为根交换机，SWD的E0/1为Alternate Port，处于Discarding状态。假设修改SWD的交换机优先级使优先级次序为SWD>SWA>SWB>SWC；协商机制的工作过程如下：

1. SWD立即成为根交换机，E0/1和E0/2立即成为指定端口，E0/2保持转发状态不变，E0/1向外发送一个Proposal（建议），Proposal是设置了一个标志位的RST BPDU，此BPDU中同时包含计算生成树的参数；

2. SWC收到Proposal之后，计算生成树，设置E0/1为根端口，保持转发状态，E0/2为指定端口。如果收到Proposal的端口是新的根端口，则设置所有非边缘指定端口为Discarding状态，并向外发送新的Proposal，如果所有的非根端口都需要进入Discarding状态或者是边缘端口，则直接在接收到Proposal的根端口上向外发送Agreement；本例中，SWC设置E0/2为Discarding状态并向外发送新的Proposal；

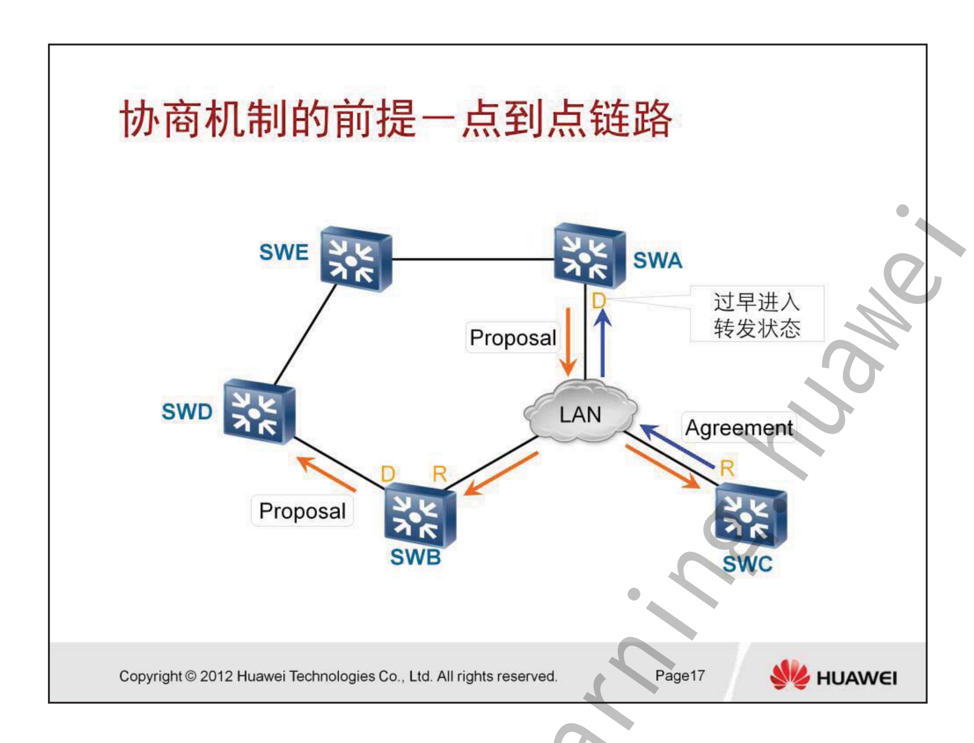
3. SWA收到Proposal之后，计算生成树，设置E0/1为预备端口，设置E0/2为根端口，如果收到Proposal的端口需要进入Discarding状态，则在该端口进入Discarding之后，向外发送一个Agreement（同意）；

更多资料下载：http://learning.huawei.com/cn

4. SWC的E0/2收到Agreement之后，立即进入转发状态，在所有非边缘指定端口收到Agreement之后，SWC在根端口上向外发送Agreement
；

5. SWD在指定端口上收到Agreement之后，立即进入转发状态。

更多资料获取：http://learning.huawei.com/cn



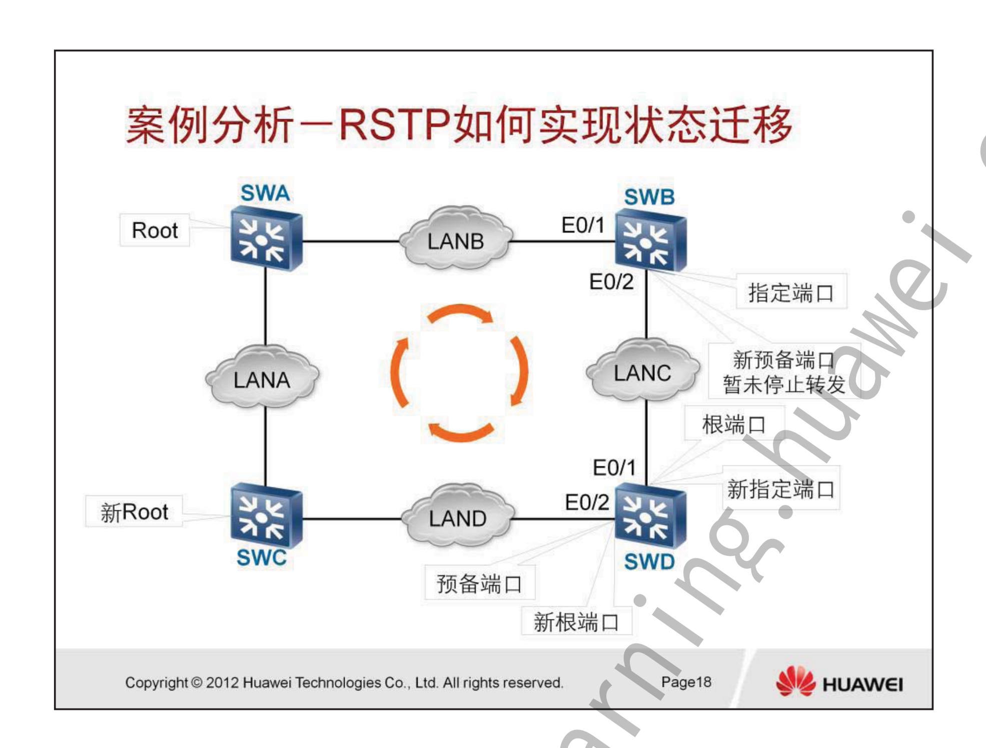
使用“Proposal – Agreement”的前提是泛洪这两种消息的链路均为点到点链路，点到点链路是指两个交换机直接相连的链路。

之所以必须使用点到点链路是因为点到多点链路有环路风险。如图所示，SWA向外发出一个Proposal之后，由于SWC是网络边缘，因此迅速返回一个Agreement，使SWA的新指定端口进入转发状态，但是此时SWB、SWD和SWE等尚未完成Proposal – Agreement的泛洪过程，因此，网络中存在环路风险。所以使用此“Proposal – Agreement”要求交换机间链路必须为点到点链路。

事实上，如果交换机间的链路没有被配置为点到点链路，泛洪过程会自动停止，需要从Discarding状态进入Forwarding状态的端口要等足够长时间（两倍Forward Delay）才能进入Forwarding状态。

实质上，“Proposal – Agreement”机制是一种在点到点链路上的“触发计算 – 确认”机制，这种“触发计算 – 确认”过程在点到点链路上泛洪，一直到达网络末端（边缘交换机，即非根端口均为边缘端口）或者预备端口（Alternate Port，处于Discarding状态，表示环路已被打断）。

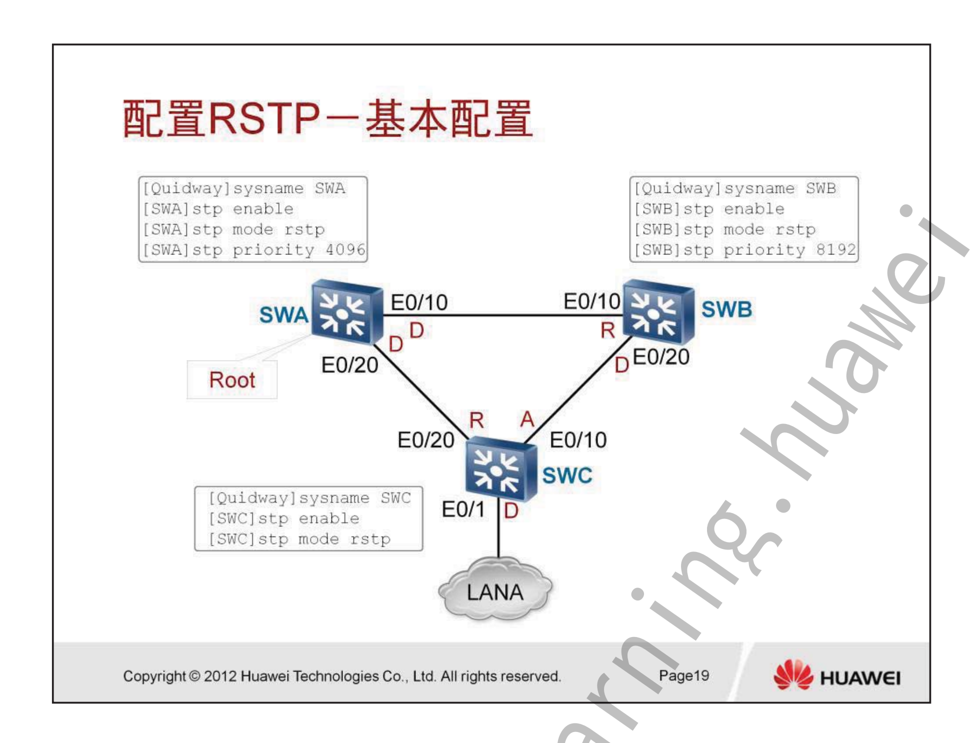
更多资料获取：http://learning.huawei.com/cn



如图所示：当SWD选举出新的根端口之后，旧的根端口成为新的指定端口，保持转发状态不变，如果在SWB的E0/2（新的预备端口）没有及时进入Discarding状态之前，SWD的E0/2（新的根端口）就进入Forwarding状态，则网络中会形成临时环路。

所以，该组网中，在SWD的E0/2成为新根端口后，立即阻断SWD的所有指定端口，E0/1作为新的指定端口被置为Discarding状态，然后新的根端口E0/2立即进入转发态。被置为Discarding的指定端口E0/1采用Proposal – Agreement机制快速变为Forwarding状态。

更多资料获取：http://learning.huawei.com/cn



本例中，要求通过修改交换机优先级使SWA成为根交换机。

配置SWA的交换机优先级为4096；配置SWB的交换机优先级为8192；保持SWC的交换机优先级为默认值（32768），使SWA成为根交换机，使各端口的角色如图所示。

stp { enable | disable }

stp命令用来启动或关闭交换机全局或端口的STP功能，缺省情况下，交换机上的STP功能处于启动状态。

stp mode { stp | rstp | mstp }

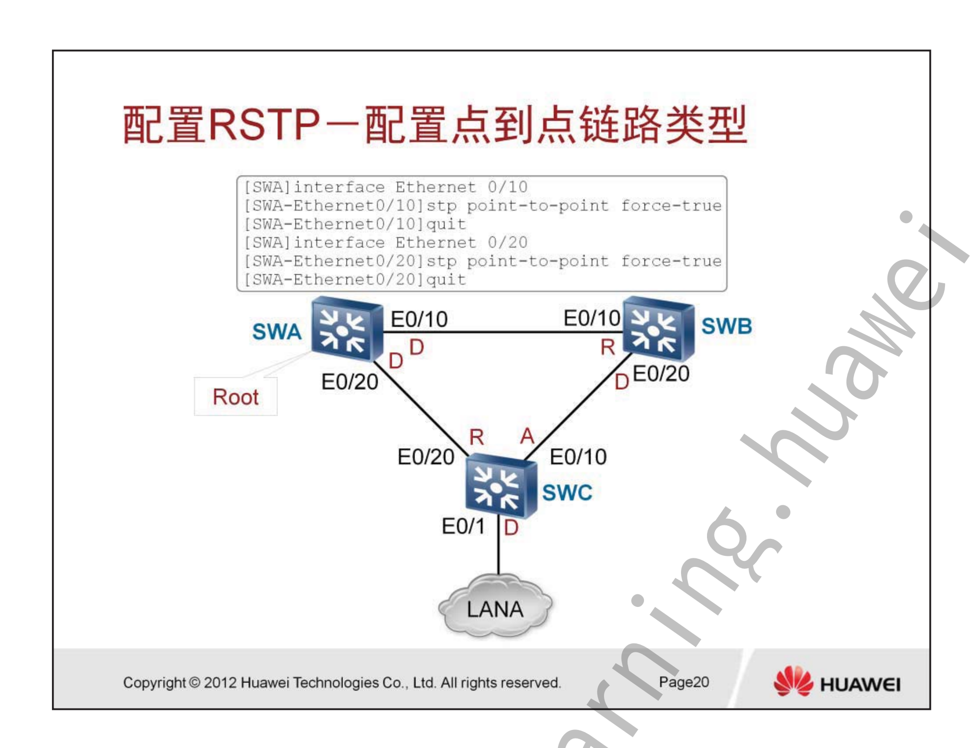
stp mode命令用来设定交换机的STP运行模式，缺省情况下，交换机的运行模式为MSTP模式。

stp priority priority

priority：交换机的优先级，取值0～61440，步长为4096，即交换机可以设置16个优先级取值，如0、4096、8192等。

stp priority命令用来配置交换机的优先级，缺省情况下，交换机优先级取值为32768。

更多资料获取：http://learning.huawei.com/cn



为了提高RSTP的收敛速度，配置交换机和交换机相连的链路为点到点链路。

stp point-to-point { force-true | force-false | auto }

force-true：用来标识与当前以太网端口相连的链路是点到点链路。

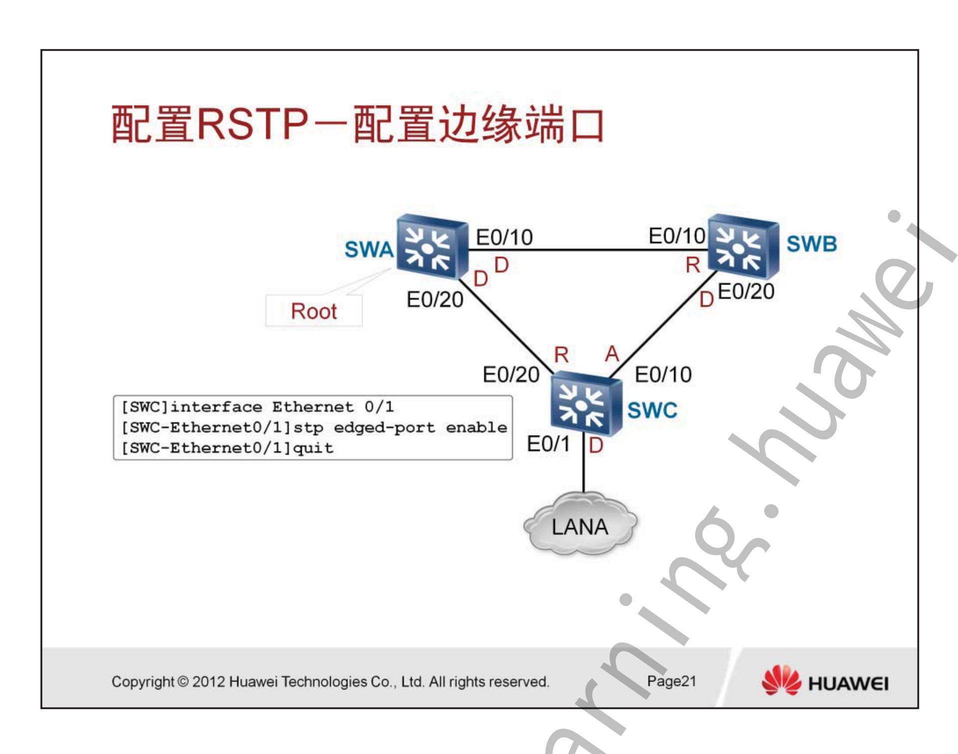
force-false：用来标识与当前以太网端口相连的链路不是点到点链路。

auto：采用自动方式检测与该以太网端口相连的链路是否是点到点链路。

缺省为auto，如果当前以太网端口工作在全双工模式，则当前端口相连的链路就被认为是点到点链路。

SWB和SWC的配置方法类同。

更多资料获取：http://learning.huawei.com/cn



可以通过命令将那些所连接的网段上没有其他交换机的端口配置成边缘端口，这样可以提高收敛速度。

stp edged-port { enable | disable }

enable：用来配置当前的以太网端口为边缘端口。

disable：用来配置当前的以太网端口为非边缘端口。

缺省情况下，交换机所有以太网端口均被配置为非边缘端口。

更多资料获取：http://learning.huawei.com/cn



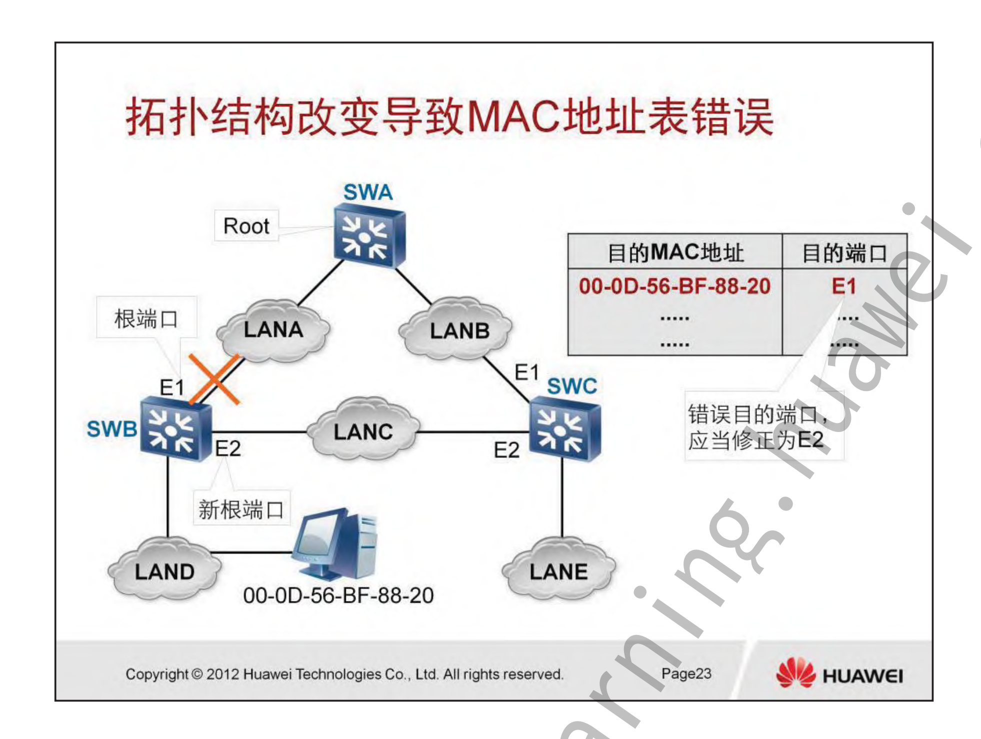
目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.





更多资料获取：http://learning.huawei.com/cn



默认情况下，MAC地址表中的动态表项生存期为300秒（5分钟）。

本例中:

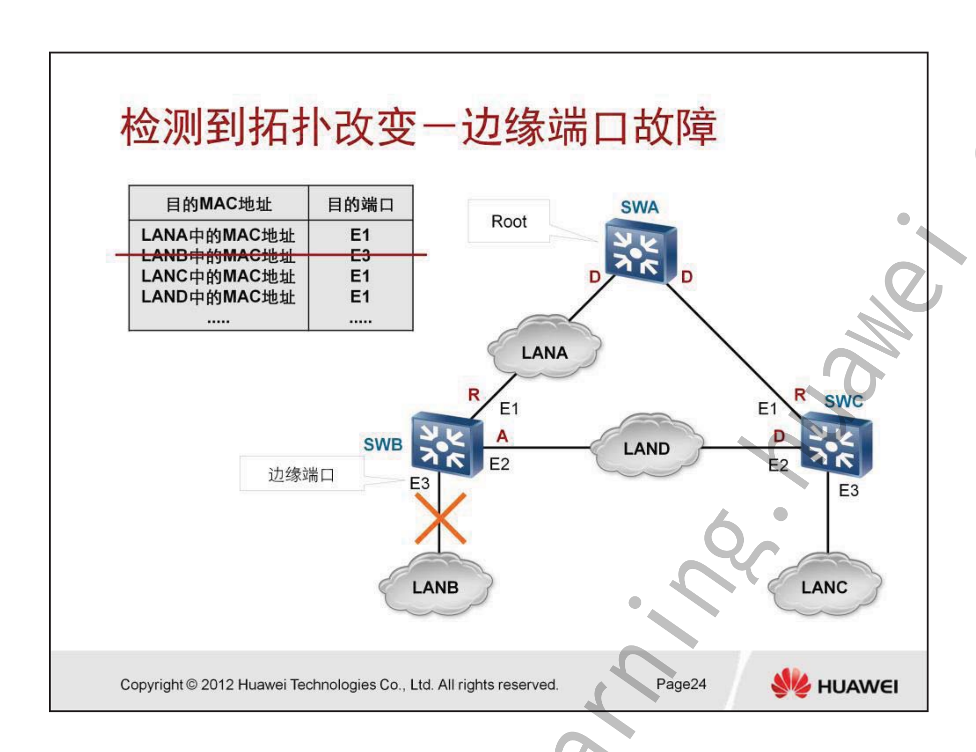
稳定拓扑下，在SWC上到达LAND某PC的目的端口应当为E1;

当SWB的E1端口断开之后，E2端口成为新的根端口，从SWC到达该PC的目的端口应当修改为E2，但是交换机不能检测到拓扑改变，导致MAC地址表错误，最长可导致5分钟的数据转发错误。

STP响应拓扑结构改变使用的是在全网泛洪拓扑改变信息，并修改MAC地址表项的生存期为一个较短值（Forward Delay，默认为15秒）。

与STP不同，RSTP响应拓扑结构改变使用的是部分删除机制。

更多资料获取：http://learning.huawei.com/cn

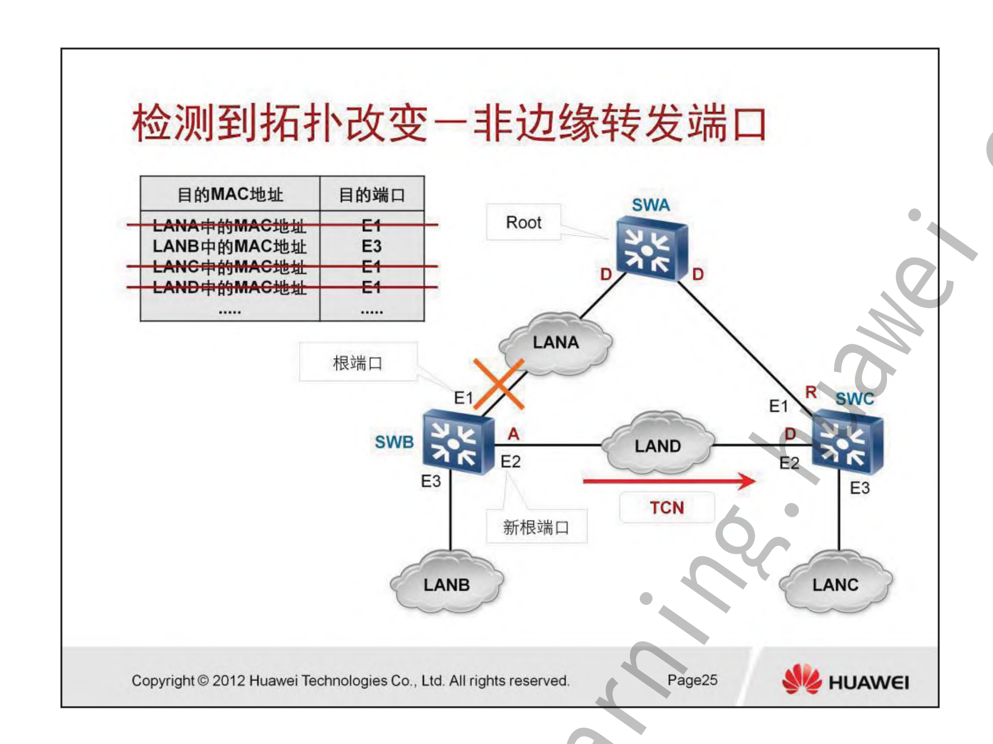


端口状态改变时，对于不同角色的端口，交换机拓扑改变处理模块的处理方式也不同。

如图所示，SWB的E3端口出现故障，SWB检测到E3端口出现故障之后，将MAC地址表中以E3端口为目的端口的所有表项删除，至此，处理过程结束。

边缘端口的任何状态改变都不会导致交换机向其它交换机发送拓扑改变信息。

更多资料获取：http://learning.huawei.com/cn



如果一个非边缘转发端口出现故障，例如图中所示，SWB的E1端口为根端口，E2为预备端口，当E1端口出现故障之后，RSTP的处理过程如下：

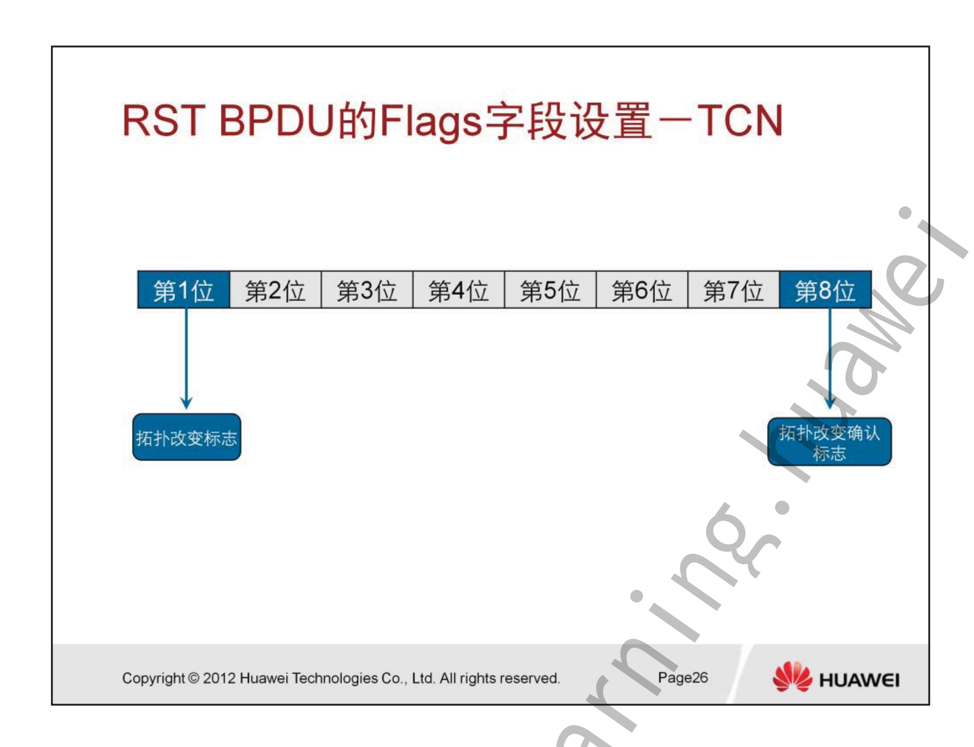
1. SWB删除MAC地址表中以E1为目的端口的端口表项；

2. 然后重新计算生成树，选举E2为新的根端口；

3. 生成树重新计算完成之后（需要进入转发状态的端口已经进入了转发状态），在所有的非边缘转发端口上向外发送拓扑改变通知（Topology Change Notification），通知其他交换机网络拓扑结构发生了改变。

当一个非边缘端口从Discarding状态进入Forwarding状态的时候，也被认为是网络拓扑结构发生了改变，交换机也会在所有处于转发状态的非边缘端口上（包括新进入Forwarding状态的端口）发送拓扑改变通知，并非只有端口故障才被认为是拓扑改变事件。

更多资料获取：http://learning.huawei.com/cn



对于RST BPDU中的Flags字段，第一位和第八位与STP使用的配置BPDU中的设置是一致的，第一位是拓扑改变标志，第八位是拓扑改变确认标志。

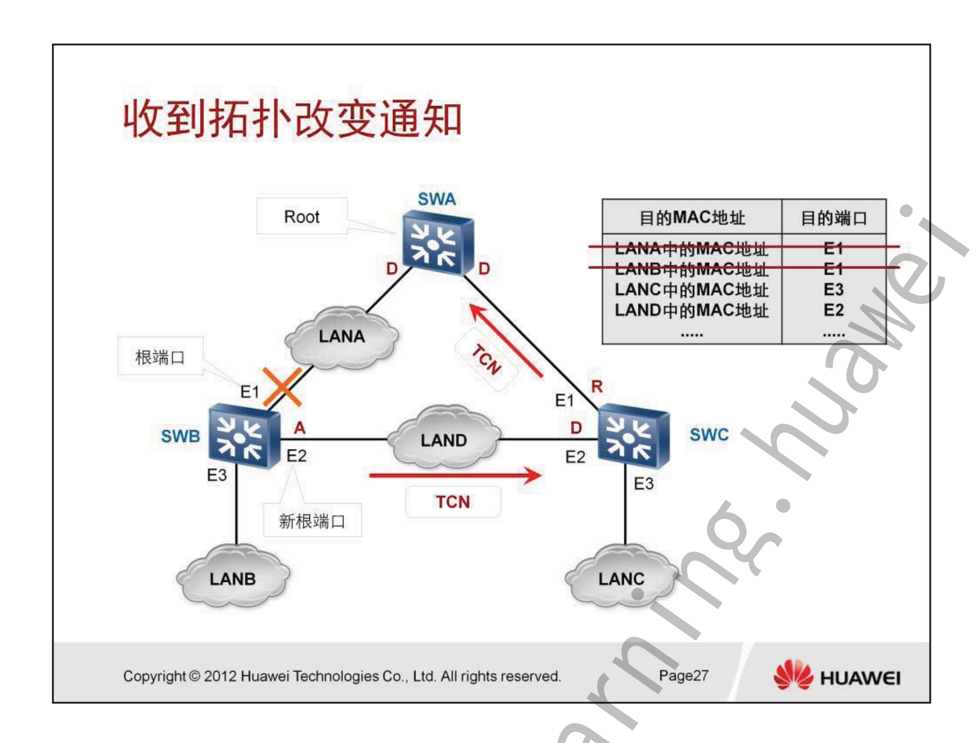
在纯RSTP环境中，RSTP交换机使用设置了拓扑改变标志的RST BPDU表示拓扑改变通知，通知其他RSTP交换机网络拓扑发生了变化。

在STP兼容环境中，RSTP交换机使用STP中的拓扑改变通知BPDU做为拓扑改变通知，通知其他STP交换机网络拓扑发生了变化。

在RSTP中，Flags字段的第八位（拓扑改变确认标志）没有使用。

在使用设置了拓扑改变标志的RST BPDU做为拓扑改变通知时，TCN定时器为Hello Time加1秒，默认为3秒。因此，当需要在纯RSTP环境中发送TCN时，一个交换机通常只发送两个TCN。

更多资料获取：http://learning.huawei.com/cn



当一个RSTP交换机从一个处于Forwarding状态的端口上收到TCN之后，处理过程如下:

1. 由于TCN也是一个RST BPDU，其中包含计算生成树的参数等信息，因此收到一个TCN之后，首先重新计算生成树（如果需要的话）；

2. 计算完成并且端口状态转换完成之后，删除以Discarding状态的端口为目的端口的表项；如果TCN的接收端口为Forwarding状态，则保留以该接收端口为目的端口的表项，保留以边缘指定端口为目的端口的表项；删除其他以Forwarding状态的端口为目的端口的表项；

3. 在既不是边缘指定端口也不是TCN接收端口并且处于Forwarding状态的端口上继续泛洪TCN。

更多资料获取：http://learning.huawei.com/cn



问 题

RSTP的端口角色有那些?

指定端口上的快速迁移机制前提是什么?

RSTP使用几种拓扑改变信息?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page28



答案:

RSTP的端口角色有那些?

RSTP的端口角色有根端口、指定端口、预备端口和备份端口四种，其中根端口和指定端口处于Forwarding状态。

指定端口上的快速迁移机制前提是什么?

指定端口所连接的链路为点到点链路。

RSTP使用几种拓扑改变信息?

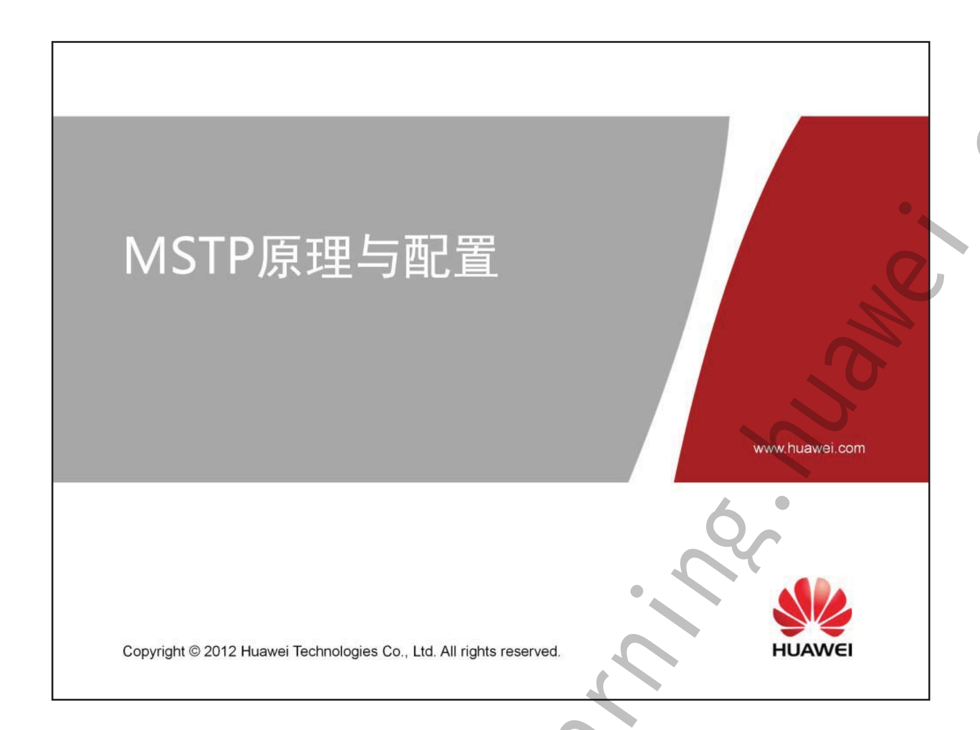
RSTP只使用一种拓扑改变信息，即设置了拓扑改变标志的RST BPDU。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

本课程介绍MSTP（多生成树协议）的原理与配置。

MSTP用于解决启用VLAN的交换网络中的环路问题。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1



本课程介绍MSTP（多生成树协议）的原理与配置。

MSTP：Multiple Spanning Tree Protocol。

为解决启用VLAN的交换网络中单个生成树的弊端，IEEE开发了MSTP，2005年版本的IEEE802.1S为MSTP当前的标准文档。

更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 描述MSTP的基本概念
- 描述MSTP高级配置

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page2



学习完此课程，您将会：

描述MSTP的基本概念，包括MST域内计算的基本概念，以及MSTP的基本配置等内容；

描述MST区域间路径的计算过程，理解跨MST区域的路径选择规则，以及相关的概念；

描述MSTP高级配置，包括针对不同交换机角色的各种保护措施等。

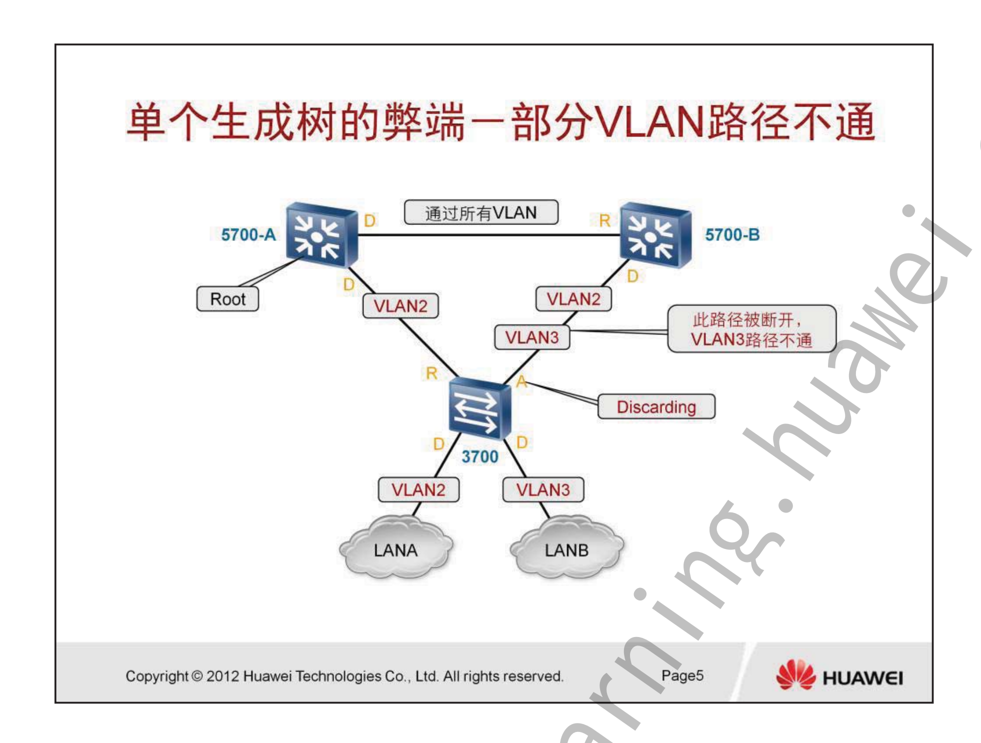
更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

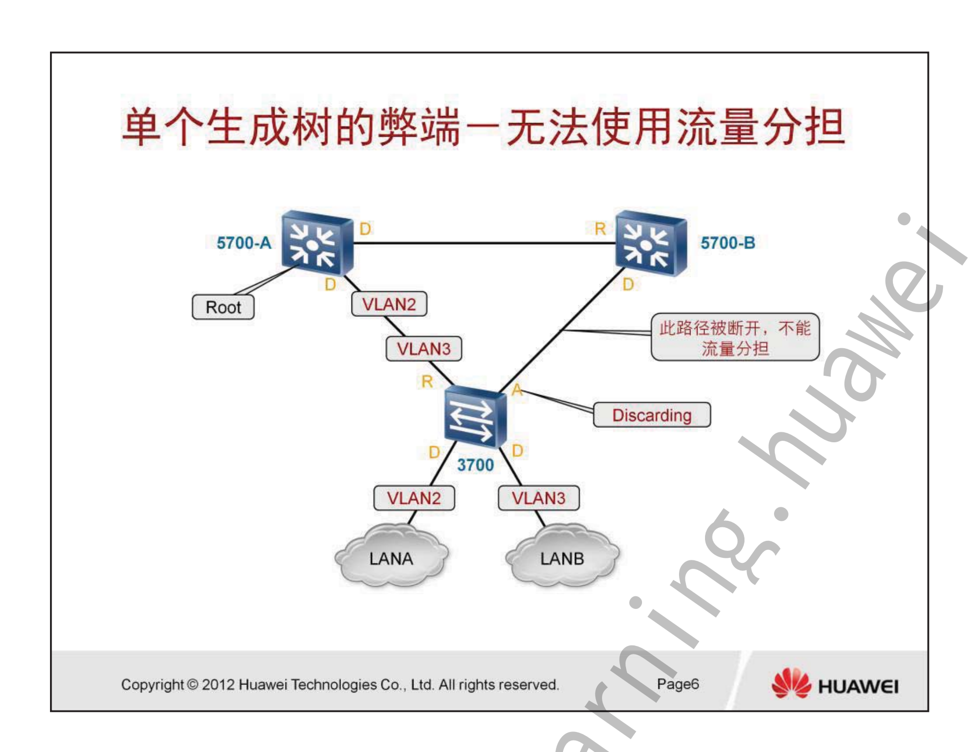


如图所示，使用一台3700连接终端网段，上行使用两条链路连接两台5700。

配置VLAN2通过两条链路上行，配置VLAN3只通过一条链路上行。

为了解决VLAN2的环路，需要运行生成树，在运行单个生成树的情况下，假设3700与5700-B相连的端口成为预备端口，进入Discarding状态。此时，VLAN3的路径被断开，就无法上行到5700-B。

更多资料获取：http://learning.huawei.com/cn



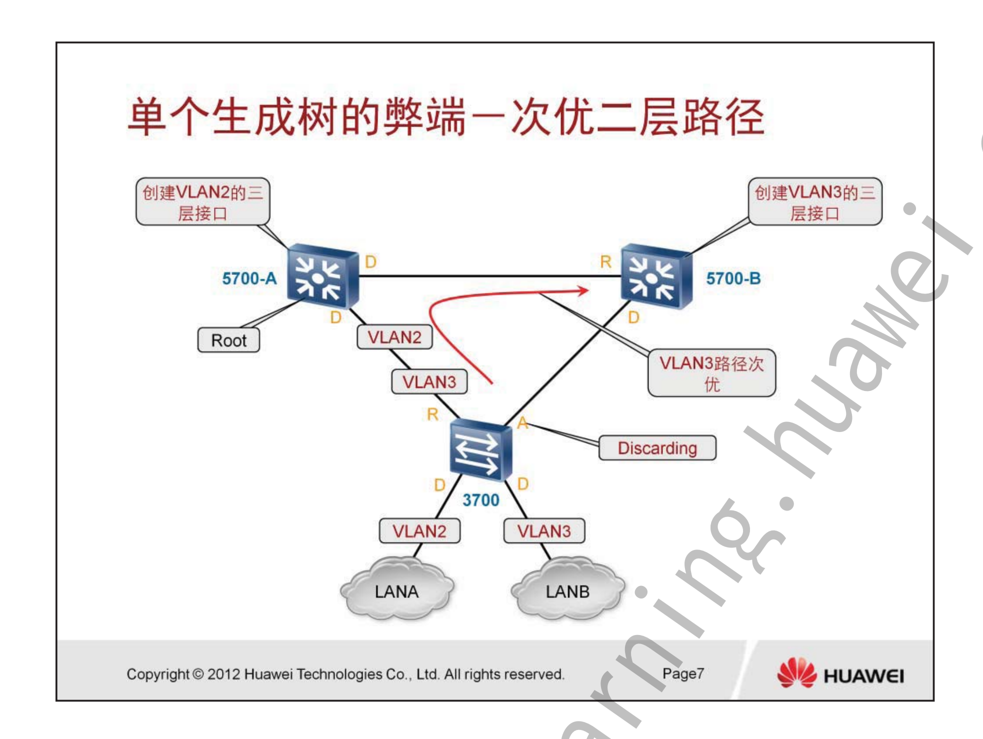
如图所示，使用一台3700连接终端网段，上行使用两条链路连接两台5700做双机热备份，并实现流量分担。

为了实现双机热备份，需要在3700上配置两条上行链路为Trunk链路，配置两条链路上都允许通过所有VLAN，两台5700之间的链路也配置为Trunk链路，允许通过所有VLAN。

将VLAN2的三层接口配置在5700-A上，将VLAN3的三层接口配置在5700-B上。

我们希望VLAN2和VLAN3分别使用不同的链路上行到相应的三层接口，可是如果网络中只有一个生成树，3700和两台5700所形成的环路就会被打开，例如，3700连接到5700-B的端口成为预备端口（Alternate Port）并处于Discarding状态，则VLAN2和VLAN3的数据都只能通过一条上行链路上行到5700-A，不能实现流量分担。

更多资料获取：http://learning.huawei.com/cn

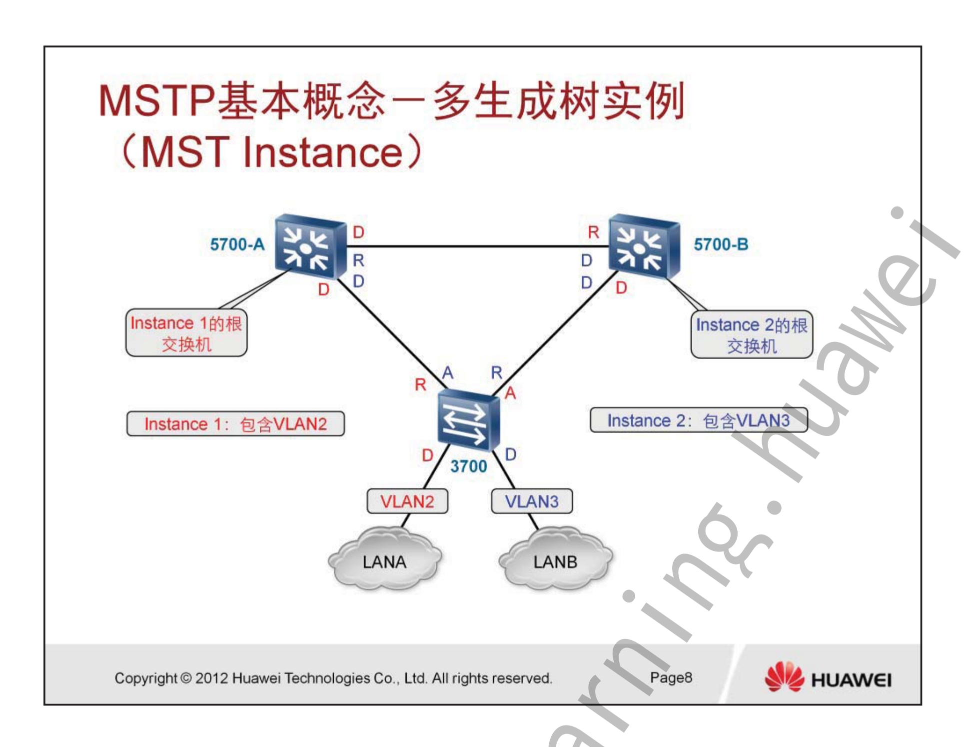


如图所示，3700和两台5700相连的链路配置为Trunk链路，允许通过所有VLAN，两台5700之间的链路也配置为Trunk链路，允许通过所有VLAN。

运行单个生成树之后，环路被打开，VLAN2和VLAN3都直接上行到5700-A。

在5700-A上配置VLAN2的三层接口，在5700-B上配置VLAN3的三层接口，则VLAN3到达三层接口的路径就是次优的，最优的路径应当是直接上行到5700-B。

更多资料获取：http://learning.huawei.com/cn



MSTP允许将一个或多个VLAN映射到一个多生成树实例（MST Instance）上，MSTP为每个MST Instance单独计算根交换机，单独设置端口状态，即在网络中计算多个生成树。

每个MST Instance都使用单独的RSTP算法，计算单独的生成树。

每个MST Intance都有一个标识（MSTID），MSTID是一个两字节的整数。

VRP平台支持16个MST Instance，MSTID取值范围是0～15，默认所有VLAN映射到MST Instance 0。

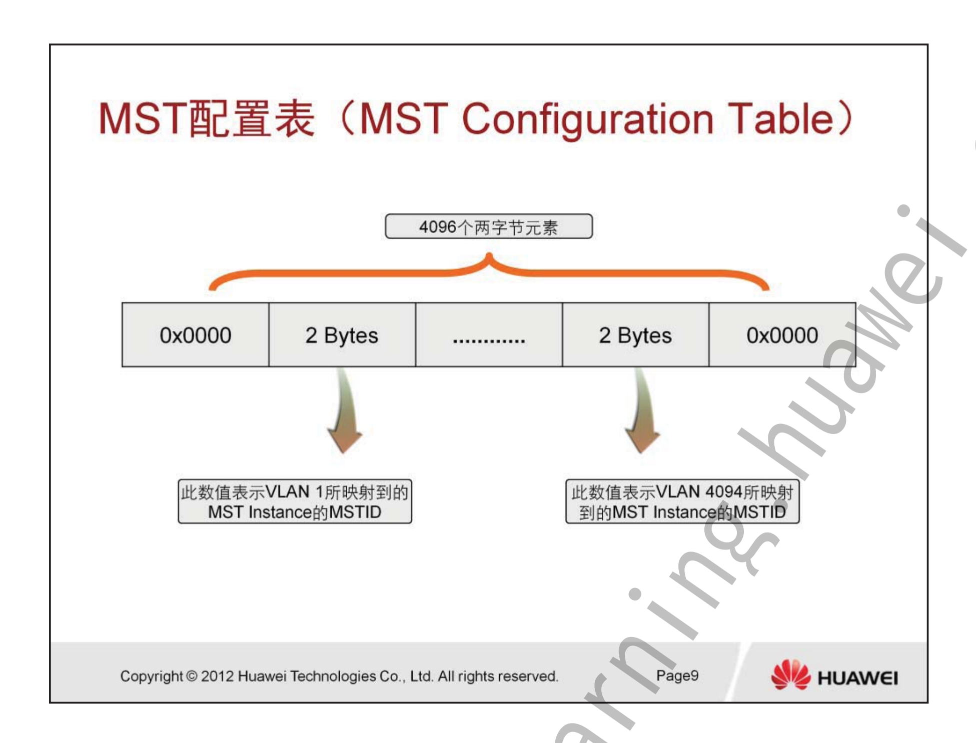
本例中，在网络中配置两个MST Instance，VLAN2映射到MST Instance 1，VLAN3映射到MST Instance 2。

通过修改交换机上不同MST Instance的交换机优先级，可以将不同的交换机设置成不同MST Instance的根交换机。

本例中，设置5700-A为MST Instance 1的根交换机；设置5700-B为MST Instance 2的根交换机。

启用了多个MST Instance之后，可以看出，VLAN2的数据直接上行到5700-A，VLAN3的数据直接上行到5700-B。如此，单生成树的弊端：流量分担，某些VLAN路径不可达，二层次优路径等问题都可以得到解决。

更多资料获取：http://learning.huawei.com/cn

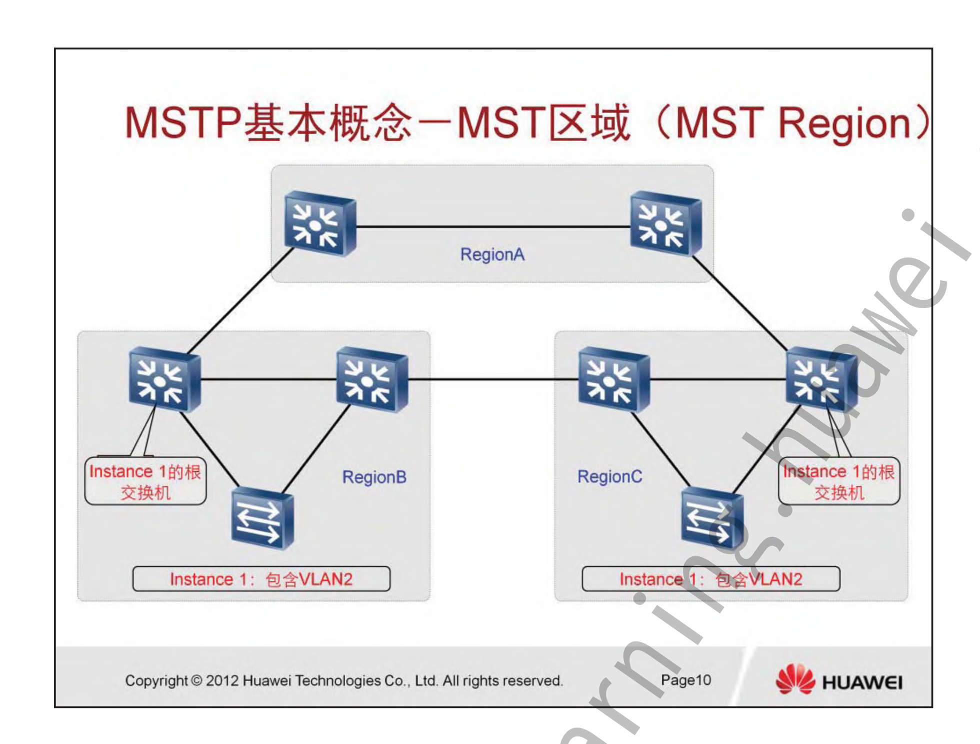


为了在交换机上标识VLAN和MST Instance的映射关系，交换机维护一个MST配置表（MST Configuration Table）。

MST配置表的结构是4096个连续的两字节元素组，代表4096个VLAN，第一个元素和最后一个元素设置为全0；第二个元素表示VLAN 1映射到的MST Instance的MSTID，第三个元素表示VLAN 2映射到的MST Instance的MSTID，依此类推，倒数第二个元素（第4095个元素）表示VLAN 4094映射到的MST Instance的MSTID。

交换机初始化时，此表格所有字段设置为全0，表示所有VLAN映射到Instance 0。

更多资料获取：http://learning.huawei.com/cn



MSTP允许一组相邻的交换机组成一个MST区域（MST Region）。同一个区域的交换机有着相同的VLAN到MST Instance的映射关系。

除了Instance 0之外（后续介绍），每个区域的MST Instance都独立计算生成树，不管是否包含相同的VLAN，不管VLAN是否通过区域间链路，区域间的生成树计算互不影响。

更多资料获取：http://learning.huawei.com/cn

MST配置标识（MST Configuration Identifier）

1 Byte	Configuration Identifier Format Selector	0x00
32 Bytes	Configuration Name	区域名称
2 Bytes	Revision Level	修订级别
16 Bytes	Configuration Digest	MST配置表摘要

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page11



交换机通过MST配置标识（MST Configuration Identifier）来标识自己所在的区域。

MST配置标识被封装在交换机相互发送的BPDU中，如图所示，MST配置标识的数据结构包括四部分，只有四部分设置都相同的相邻交换机才被认为是在同一个区域中。

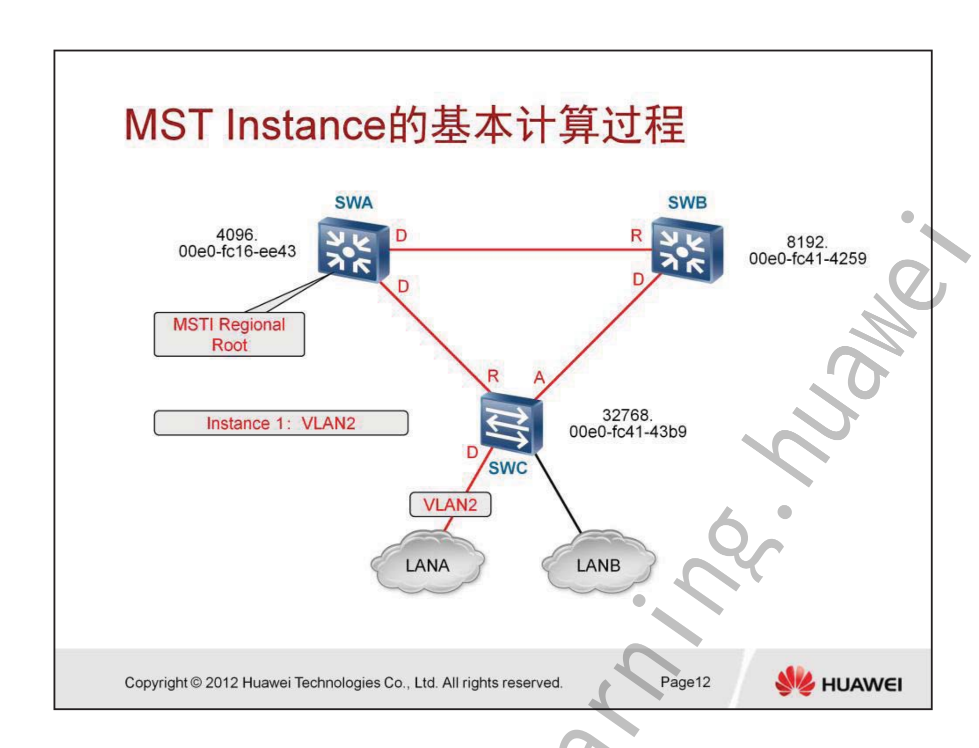
Configuration Identifier Format Selector：配置标识格式选择符，长度为一个字节，固定设置为0。

Configuration Name：配置名称，也就是交换机的MST域名，长度为32字节。每个交换机都配置一个MST域名，默认为交换机的MAC地址。

Configuration Digest：配置摘要，长度为16字节。相同区域的交换机应当维护相同的VLAN到MST Instance的映射表，可是MST配置表太大（8192字节），不适合在交换机之间相互发送。此字段是使用MD5算法从MST配置表中算出的摘要信息。

Revision Level：修订级别，长度为两个字节，默认取值为全0。由于Configuration Digest是MST配置表的摘要信息，因此有很小的可能会出现MST配置表不同但摘要信息却相同的情况，这会导致本来不在同一区域的交换机被认为在同一区域中，此字段是一个额外的标识字段，建议不同的区域使用不同的数值，以消除上述可能产生错误的情况。

更多资料获取 http://learning.huawei.com/cn



本文使用的术语MST Instance指MST Instance 1到MST Instance 15，也就是非0的MST Instance，Instance 0的概念和相关计算后续章节介绍。

每个MST Instance的基本计算过程也就是RSTP的计算过程，只是在术语上有些差别：

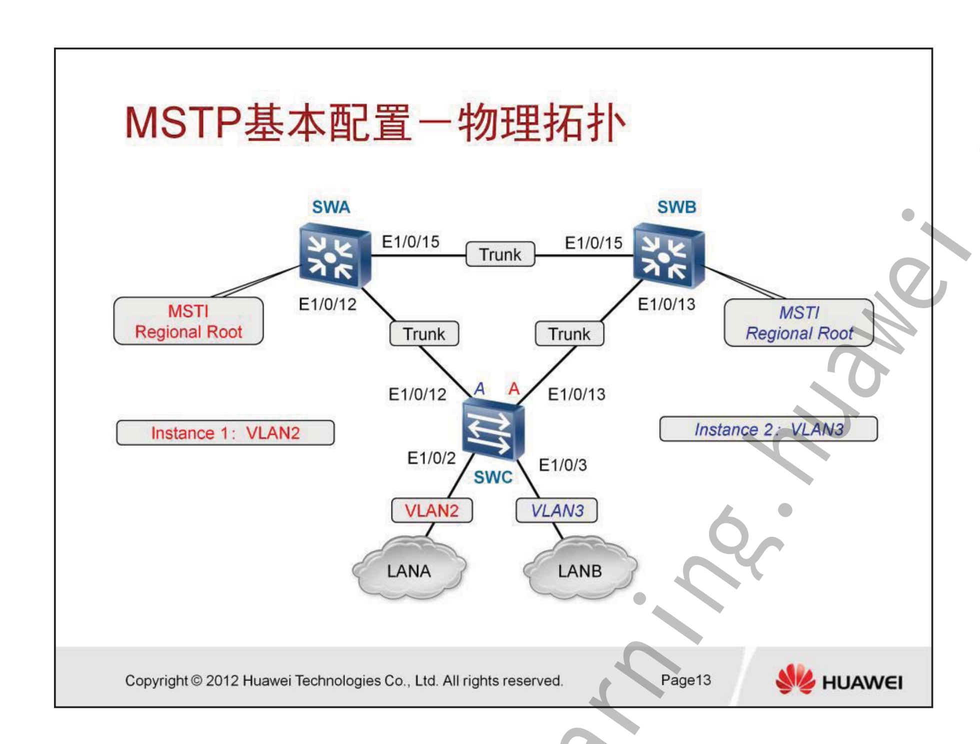
1. 计算过程首先选择此MST Instance的MSTI Regional Root（MSTI区域根交换机），相当于RSTP中的根交换机。选举的依据是各交换机配置在该MST Instance中的交换机标识，如同RSTP，此交换机标识由交换机优先级和MAC地址两部分组成，数值越小越优先。

2. 此MST Instance的非根交换机选举一个根端口，根端口为该交换机提供到达此MST Instance的MSTI Regional Root的最优路径。选举的依据为Internal Root Path Cost（内部根路径开销），表示一个交换机到达相关MSTI Regional Root的MST区域内部开销，如果多个端口提供的路径开销相同，则按顺序比较上行交换机标识、所连接上行交换机端口的端口标识以及接收端口的端口标识。

3. 每个网段的指定端口为所连接网段提供到达相关MSTI Regional Root的最优路径。

4. 预备端口和备份端口的选择依据和RSTP相同。

更多资料获取：http://learning.huawei.com/cn



如图所示，所有交换机之间的链路配置成Trunk链路，并配置允许通过所有VLAN。

将网络中的三台交换机都配置在一个MST区域中，Configration Name（MST域名）为“RegionA”，Revision Level为“1”，在区域中新建两个MST Instance，Instance 1包含VLAN 2，Instance 2包含VLAN 3。

通过修改交换机在不同Instance中的优先级，使SWA成为Instance 1的根交换机，使SWC的E1/0/13成为Instance 1的预备端口（Alternate Port）；使SWB成为Instance 2的根交换机，使SWC的E1/0/12成为Instance 2的预备端口（Alternate Port）。

如此配置，可以使VLAN 2和VLAN 3沿不同的链路上行，实现流量分担的目的，并使SWC的两条上行链路相互备份。

更多资料获取：http://learning.huawei.com/cn

MSTP基本配置－配置SWA的MST域参数

```
[SWA]stp enable
[SWA]stp mode mstp
[SWA]stp region-configuration
[SWA-mst-region]region-name RegionA
[SWA-mst-region]revision-level 1
[SWA-mst-region]instance 1 vlan 2
[SWA-mst-region]instance 2 vlan 3
[SWA-mst-region]active region-configuration
[SWA-mst-region]quit
[SWA]stp instance 1 priority 4096
[SWA]stp instance 2 priority 8192
[SWA]
```

说明：SWB、SWC配置MST域参数类似，不再赘述。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page14



stp { enable | disable }

stp命令用来启动或关闭交换机全局或端口的MSTP特性。缺省情况下，交换机上的MSTP特性处于启动状态。

stp mode { stp | rstp | mstp }

stp mode命令用来设置交换机的MSTP工作模式。缺省值为MSTP模式。

stp region-configuration命令用来进入MST域视图。

region-name name

name：交换机的MST域名（Configuration Name），为1～32位字符串。缺省情况下，交换机的MST域名为交换机的MAC地址。

revision-level level

level：MSTP修订级别，取值范围为0～65535。缺省情况下，MSTP修订级别取值为0。

instance instance-id vlan vlan-list

instance命令用来将所指定的VLAN列表映射到指定的MST Instance上。缺省所有VLAN映射到Instance 0上。

active region-configuration

active region-configuration命令用来激活MST域的配置。

stp [instance instance-id] priority priority

stp priority命令用来配置交换机在指定MST Instance上的优先级，数值为4096的整数倍。每个Instance的默认Priority为32768。

更多资料获取：http://learning.huawei.com/cn

MSTP基本配置－设置RSTP点到点链路和边缘端口

```
[SWA]interface Ethernet 1/0/12
[SWA-Ethernet1/0/12]stp point-to-point force-true
[SWA]interface Ethernet 1/0/15
[SWA-Ethernet1/0/15]stp point-to-point force-true
```

```
[SWB]interface Ethernet 1/0/13
[SWB-Ethernet1/0/13]stp point-to-point force-true
[SWB]interface Ethernet 1/0/15
[SWB-Ethernet1/0/15]stp point-to-point force-true
```

```
[SWC]interface Ethernet 1/0/12
[SWC-Ethernet1/0/12]stp point-to-point force-true
[SWC]interface Ethernet 1/0/13
[SWC-Ethernet1/0/13]stp point-to-point force-true
[SWC]interface Ethernet 1/0/2
[SWC-Ethernet1/0/2]stp edged-port enable
[SWC]interface Ethernet 1/0/3
[SWC-Ethernet1/0/3]stp edged-port enable
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page16



每个MST Instance都使用单独的RSTP算法计算生成树，RSTP的快速收敛机制对每个MST Instance都是有效的。

如同RSTP的配置一样，此处，设置交换机之间的链路为点到点链路，设置SWC的E1/0/2和E1/0/3两个端口为边缘端口。

stp point-to-point { force-true | force-false | auto }

force-true：用来标识与当前以太网端口相连的链路是点到点链路。

force-false：用来标识与当前以太网端口相连的链路不是点到点链路。

auto：采用自动方式检测与该以太网端口相连的链路是否是点到点链路。

缺省为auto，当检测到端口工作在全双工模式下的时候，认为端口所连接的链路是点到点链路，当检测到端口工作在半双工模式下的时候，认为端口所连接的链路不是点到点链路。此处使用强制命令设置为点到点链路。

stp edged-port { enable | disable }

stp edged-port enable命令用来将当前的以太网端口配置为边缘端口。

stp edged-port disable命令用来将当前的以太网端口配置为非边缘端口。缺省情况下，交换机所有以太网端口均被配置为非边缘端口。

更多资料获取：http://learning.huawei.com/cn



MSTP基本配置－验证MSTP基本信息

```
[SWA]display stp brief
 MSTID      Port                Role  STP State      Protection
   0        Ethernet1/0/12      DESI  FORWARDING       NONE
   0        Ethernet1/0/15      ROOT  FORWARDING       NONE
   1        Ethernet1/0/12      DESI  FORWARDING       NONE
   1        Ethernet1/0/15      DESI  FORWARDING       NONE
   2        Ethernet1/0/12      DESI  FORWARDING       NONE
   2        Ethernet1/0/15      ROOT  FORWARDING       NONE
```

```
[SWB]display stp brief
 MSTID      Port                Role  STP State      Protection
   0        Ethernet1/0/13      DESI  FORWARDING       NONE
   0        Ethernet1/0/15      DESI  FORWARDING       NONE
   1        Ethernet1/0/13      DESI  FORWARDING       NONE
   1        Ethernet1/0/15      ROOT  FORWARDING       NONE
   2        Ethernet1/0/13      DESI  FORWARDING       NONE
   2        Ethernet1/0/15      DESI  FORWARDING       NONE
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page17 

暂时忽略Instance 0的信息（后续介绍）。

在SWA上，Instance 1的两个端口都是指定端口（Designated Port），表明SWA是Instance 1的根交换机；

在SWB上，Instance 2的两个端口都是指定端口（Designated Port），表明SWB是Instance 2的根交换机。

更多资料获取：http://learning.huawei.com/cn

MSTP基本配置－验证MSTP基本信息

```
[SWC]display stp brief
 MSTID      Port                 Role  STP State     Protection
   0        Ethernet1/0/2        DESI  FORWARDING      NONE
   0        Ethernet1/0/3        DESI  FORWARDING      NONE
   0        Ethernet1/0/12       ALTE  DISCARDING      NONE
   0        Ethernet1/0/13       ROOT  FORWARDING      NONE
   1        Ethernet1/0/2        DESI  FORWARDING      NONE
   1        Ethernet1/0/12       ROOT  FORWARDING      NONE
   1        Ethernet1/0/13       ALTE  DISCARDING      NONE
   2        Ethernet1/0/3        DESI  FORWARDING      NONE
   2        Ethernet1/0/12       ALTE  DISCARDING      NONE
   2        Ethernet1/0/13       ROOT  FORWARDING      NONE
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page18



暂时忽略Instance 0的信息（后续介绍）。

如图所示，在Instance 1中，SWC的E1/0/13成为预备端口（Alternate Port），处于Discarding状态；在Instance 2中，SWC的E1/0/12成为预备端口（Alternate Port），处于Discarding状态。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

MSTP工作模式

```
[SWB]stp mode stp
[SWB]stp mode rstp
[SWB]stp mode mstp
```

工作模式	描述
STP	只能和STP交换机交互，只能在端口上收发配置BPDU。
RSTP	运行RSTP，如果检测到端口相邻的交换机运行在STP模式下，则运行STP。
MSTP	运行MSTP，如果检测到端口相邻的交换机运行在RSTP模式下，则运行RSTP，如果检测到端口相邻的交换机运行在STP模式，则运行STP。

```
[SWB]stp mcheck
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page20 

工作模式可以在全局模式下配置，也可以在端口模式下配置。

三种工作模式总的原则就是向下兼容，MSTP兼容RSTP，RSTP兼容STP。

如果MSTP交换机的端口上曾经连接有STP/RSTP交换机，则该端口被迁移到STP/RSTP兼容工作模式。如果STP/RSTP交换机被关机或移走，该端口无法自动迁移到MSTP模式下工作。此时如果在端口上执行mcheck操作，则该端口会重新迁移到MSTP模式下工作。

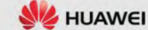
stp mcheck命令用来在当前端口执行mcheck操作。

更多资料获取：http://learning.huawei.com/cn

设置交换机为主用/备用根交换机

```
[SWA]stp instance 0 root primary
[SWA]display stp instance 0
-------[CIST Global Info][Mode MSTP]-------
CIST Bridge         :0.000f-e212-f8e1
Bridge Times        :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC      :0.000f-e212-f8e1 / 0
CIST RegRoot/IRPC   :0.000f-e212-f8e1 / 0
CIST RootPortId     :0.0
CIST Root Type      :PRIMARY root
```

```
[SWB]stp instance 0 root secondary
[SWB]display stp instance 0
-------[CIST Global Info][Mode MSTP]-------
CIST Bridge         :4096.000f-e212-f890
Bridge Times        :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC      :0.000f-e212-f8e1 / 199999
CIST RegRoot/IRPC   :4096.000f-e212-f890 / 0
CIST RootPortId     :128.13
CIST Root Type      :SECONDARY root
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page21 

VRP平台支持将交换机配置为主根交换机或者备用根交换机，避免了手动配置优先级的麻烦。

stp [instance instance-id] root primary

此命令自动修改所指定的Instance的优先级为0，使此交换机成为所指定的Instance的主用根交换机。

stp [instance instance-id] root secondary

此命令自动修改所指定的Instance的优先级为4096，使此交换机成为所指定的Instance的备用根交换机，当主用根交换机故障之后，可以立即成为主用根交换机。

这两个命令的目的是为了提供另一种修改STP优先级的方式。

更多资料获取：http://learning.huawei.com/cn

配置MSTP最大跳数

```
[SWA]stp max-hops 30
[SWA]display stp
-------[CIST Global Info][Mode MSTP]-------
CIST Bridge          :0.000f-e212-f8e1
Bridge Times         :Hello 2s MaxAge 20s FwDly 15s MaxHop 30
CIST Root/ERPC       :0.000f-e212-f8e1 / 0
CIST RegRoot/IRPC    :0.000f-e212-f8e1 / 0
CIST RootPortId      :0.0
BPDU-Protection      :disabled
CIST Root Type       :PRIMARY root
TC or TCN received   :3
Time since last TC   :0 days 1h:23m:36s
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page22 HUAWEI

MSTP支持最大跳数的概念，在MST BPDU中，有一个剩余跳数（CIST Remaining Hops）字段，类似于IP报文中TTL值的字段。

当一个MST BPDU从MST域根交换机发出时，此字段设置为此MST域根交换机上配置的最大跳数（Max Hops）参数，每一个非根交换机在生成自己的MST BPDU并向下游发送时，会将此MST BPDU中的剩余跳数字段设置成从上游交换机接收到的MST BPDU的剩余跳数减一。

当交换机收到剩余跳数为0的MST BPDU时，会将此MST BPDU丢弃，使处于最大跳数之外的交换机不能参与生成树计算，限制MST域的规模。

stp max-hops命令用来在交换机上设置MST域的最大跳数，此命令只在MST域内的域根交换机上起作用。VRP平台支持1－40跳，默认为20跳。

更多资料获取：http://learning.huawei.com/cn

调整时间参数

```
[SWA]stp timer ?
  forward-delay  Specify forward delay
  hello          Specify hello time interval
  max-age        Specify max age
```

```
[SWA]stp timer-factor ?
  INTEGER<1-10>  Aged out time factor
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



stp timer forward-delay centi-seconds

centi-seconds：Forward Delay时间参数，取值范围为400～3000，单位为百分之一秒。 默认取值为15秒。

stp timer hello centi-seconds

centi-seconds：Hello Time时间参数，取值范围为100～1000，单位为百分之一秒。默认取值为2秒。

stp timer max-age centi-seconds

centi-seconds：Max Age时间参数，取值范围为600～4000，单位为百分之一秒。默认取值为20秒。

stp timer-factor number

number：用来设定超时时间，设定的数值是Hello Time的倍数，范围1～10。缺省情况下，倍数为3。表示如果在端口上3倍Hello间隔（共6秒）没有收到所连接网段的指定端口发出的BPDU，则认为指定端口发生故障，应重新计算生成树。

更多资料获取：http://learning.huawei.com/cn

网络直径与时间参数的关系

网络直径	Hello Timer	Max Age	Forward Delay
2	2s	10s	7s
3	2s	12s	9s
4	2s	14s	10s
5	2s	16s	12s
6	2s	18s	13s
7	2s	20s	15s

```
[SWA]stp bridge-diameter ?
  INTEGER<2-7>   Bridge diameter
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page24



当Max Age或者Forward Delay配置不合理的时候，会使网络中可能产生环路或者拓扑改变之后网络长时间不通。

VRP可以根据配置的网络直径和Hello间隔自动计算比较合理的Max Age和Forward Delay，表中列出了当Hello间隔为2秒的时候，根据不同的网络直径，VRP自动计算出的Max Age和Forward Delay参数。

stp bridge-diameter bridgenum

bridgenum：交换网络的网络直径，取值为2～7，缺省为7。

更多资料获取：http://learning.huawei.com/cn

配置边缘端口保护

```
[SWA]stp bpdu-protection
[SWA]
[SWA]display stp
-------[CIST Global Info][Mode MSTP]-------
CIST Bridge         :32768.000f-e212-f8e1
Bridge Times        :Hello 2s MaxAge 20s FwDly 15s MaxHop 20
CIST Root/ERPC      :32768.000f-e212-f8e1 / 0
CIST RegRoot/IRPC   :32768.000f-e212-f8e1 / 0
CIST RootPortId     :0.0
BPDU-Protection     :enabled
TC or TCN received  :0
Time since last TC  :0 days 1h:19m:5s
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page25 HUAWEI

边缘端口正常情况下是不会收到BPDU的，如果边缘端口收到了伪造的更优的BPDU会重新计算生成树，造成拓扑振荡。

stp bpdu-protection用于开启边缘端口的保护功能，保护功能开启之后，如果在边缘端口上收到了BPDU，则认为交换机受到了攻击，收到BPDU的边缘端口将自动关闭，需要网络管理员手动开启。

默认不开启边缘端口保护功能。

更多资料获取：http://learning.huawei.com/cn

配置根交换机的指定端口保护

```
[SWA] interface Ethernet 1/0/13
[SWA-Ethernet0/0/13] stp root-protection
[SWA]display stp interface Ethernet 1/0/13

----[CIST][Port13(Ethernet1/0/13)][FORWARDING]----
 Port Protocol          :enabled
 Port Role              :CIST Designated Port
 Port Priority          :128
 Port Cost(Dot1T)       :Config=auto / Active=199999
 Desg. Bridge/Port      :0.000f-e212-f8e1 / 128.13
 Port Edged             :Config=disabled / Active=disabled
 Point-to-point         :Config=auto / Active=true
 Transit Limit          :3 packets/hello-time
 Protection Type        :Root
 Num of Vlans Mapped :1
 PortTimes              :Hello 2s MaxAge 20s FwDly 15s RemHop 0
 BPDU Sent              :18
          TCN: 0, Config: 0, RST: 0, MST: 18
 BPDU Received          :0
          TCN: 0, Config: 0, RST: 0, MST: 0
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page26 HUAWEI

网络设计时一般会把CIST的根桥和备份根桥放在一个高带宽的核心域内。

但是由于维护人员的错误配置或网络中的恶意攻击，网络中的合法根桥有可能会收到优先级更高的配置消息，这样当前根桥会失去根桥的地位，引起网络拓扑结构的错误变动。这种不合法的变动，会导致原来应该通过高速链路的流量被牵引到低速链路上，导致网络拥塞。

Root保护功能可以防止这种情况的发生。对于设置了Root保护功能的端口，其在所有实例上的端口角色只能保持为指定端口。一旦这种端口上收到了更优的配置消息，该端口将被选择为非指定端口，这些端口的状态将被设置为Discarding状态，不再转发报文（相当于将与此端口相连的链路断开）。当在足够长的时间内（Max Age，默认20秒）没有收到更优的配置消息时，端口会恢复原来的正常状态，重新成为指定端口，进入转发状态。

即，该保护功能用于保护根交换机的指定端口回避攻击。

更多资料获取：http://learning.huawei.com/cn

配置环路保护功能

```
[SWB] interface Ethernet 1/0/13
[SWB-Ethernet0/0/13] stp loop-protection
[SWB]display stp interface Ethernet 1/0/13

----[CIST][Port13(Ethernet1/0/13)][FORWARDING]----
 Port Protocol          :enabled
 Port Role              :CIST Root Port
 Port Priority          :128
 Port Cost(Dot1T)       :Config=auto / Active=199999
 Desg. Bridge/Port      :0.000f-e212-f8e1 / 128.13
 Port Edged             :Config=disabled / Active=disabled
 Point-to-point         :Config=auto / Active=true
 Transit Limit          :3 packets/hello-time
 Protection Type        :Loop
 Num of Vlans Mapped    :1
 PortTimes              :Hello 2s MaxAge 20s FwDly 15s RemHop 0
 BPDU Sent              :6
          TCN: 0, Config: 0, RST: 0, MST: 6
 BPDU Received          :705
          TCN: 0, Config: 0, RST: 0, MST: 705
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page27 HUAWEI

依靠不断接收上游交换机发送的BPDU，交换机可以维持根端口和其他阻塞端口的状态。

但是由于链路拥塞或者单向链路故障，这些端口有可能会收不到上游交换机指定端口发送的BPDU。此时交换机会重新选择根端口，根端口会转变为指定端口，而阻塞端口会迁移到转发状态，从而交换网络中会产生环路。

环路保护功能会抑制这种环路的产生。在启动了环路保护功能后，如果根端口或Alternate端口长时间收不到来自上游的BPDU时，则向网管发出通知信息（如果是根端口则进入Discarding状态）。而阻塞端口则会一直保持在阻塞状态，不转发报文，从而不会在网络中形成环路。直到根端口收到BPDU，端口状态才恢复正常为Forwarding状态。

因此，环路保护功能会抑制由于链路拥塞等原因产生的环路。

更多资料获取：http://learning.huawei.com/cn

配置TC-BPDU保护功能

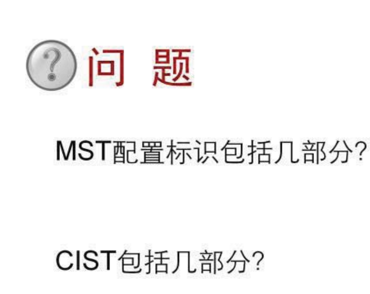
```
[SWB]tc-protection ?
  threshold   Set the threshold value
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.
Page28 HUAWEI

交换机在接收到TC-BPDU报文后，会执行MAC地址表项和ARP表项的删除操作。当有人伪造TC-BPDU报文恶意攻击交换机时，交换机短时间内会收到很多的TC-BPDU报文，频繁的删除操作会给交换机带来很大负担，给网络的稳定带来很大隐患。

开启了TC-BPDU报文攻击的保护功能后，在单位时间内，交换设备处理拓扑变化报文的次数可配置。如果在单位时间内，交换设备在收到拓扑变化报文数量大于配置的阈值，那么设备只会处理阈值指定的次数。对于其他超出阈值的拓扑变化报文，定时器到期后设备只对其统一处理一次。这样可以避免频繁的删除MAC地址表项和ARP表项，从而达到保护设备的目的。

更多资料获取：http://learning.huawei.com/cn



问 题

MST配置标识包括几部分?

CIST包括几部分?

Master端口有什么作用?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page29



答案:

MST配置标识包括几部分?

包括配置标识格式选择符，配置名称，配置摘要，修订级别四部分内容。只有四部分都一致的相邻交换机，才被认为是在同一区域内部。

CIST包括几部分?

包括CST和MST区域内的IST两部分，用于连接网络中的所有交换机和网段。

Master端口有什么作用?

用于连接MST Instance到CIST的总根。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn

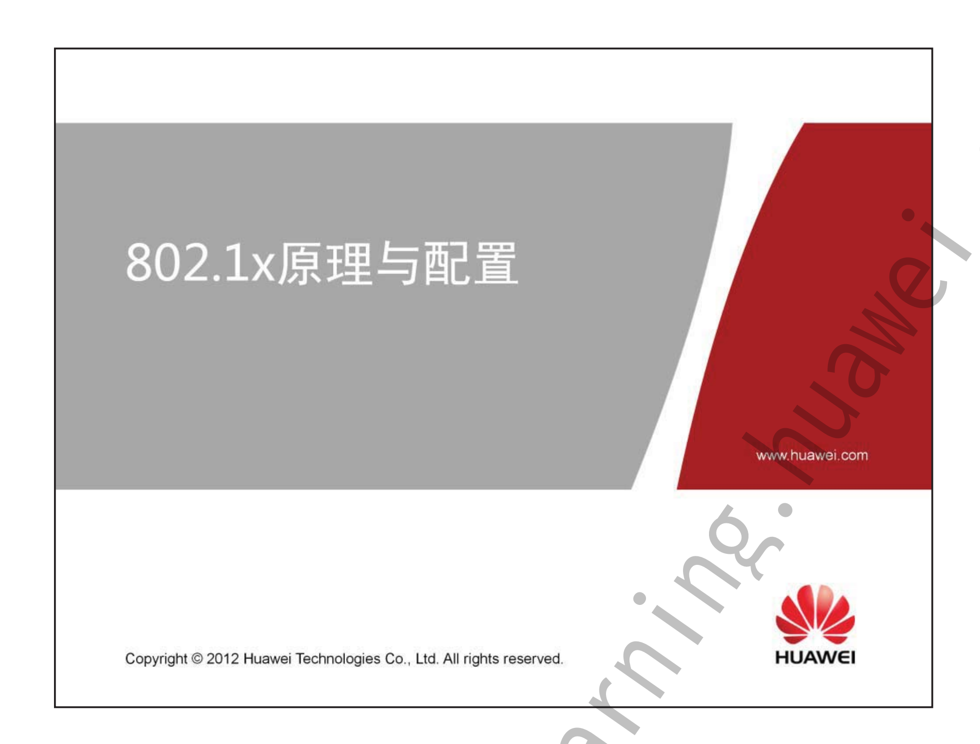
更多资料获取：http://learning.huawei.com/cn

Module 3

接入技术

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

当前，网络安全已超过对交换能力和服务质量等的需求，成为企业用户最关心的问题之一。

在企业网络中，任何一台终端的安全状态都将直接影响到整个网络的安全。而传统针对病毒的防御体系是以孤立的单点防御为主，这样的分散管理无法避免诸多的安全威胁。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 理解并解释802.1x技术原理
- 描述802.1x系统组件和工作过程
- 掌握802.1x的简单场景配置
- 掌握诊断802.1x故障的基本能力

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page2 HUAWEI

更多资料获取：http://learning.huawei.com/cn



目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page3



NAC:网络准入控制

病毒、蠕虫和间谍软件等威胁损害客户利益并使机构损失大量的金钱、生产率和机会。与此同时，移动计算的普及进一步加剧了威胁。移动用户能够从家里或公共热点连接互联网或办公室网络，常在无意中轻易地感染病毒并将其带进企业环境，进而感染网络。 网络准入控制(NAC) 进行了专门设计，可确保为访问网络资源的所有终端设备(如PC、笔记本电脑、服务器、智能电话或PDA等)提供足够保护，以增强网络安全。

更多资料获取：http://learning.huawei.com/cn

NAC技术产生背景

企业内网安全面临的主要问题是内部威胁（高达60%），而终端是威胁的主要来源：

- 终端不能及时打系统补丁
- 员工绕过防火墙访问互联网
- 员工未安装防病毒软件
- 员工忘记设置必要的口令

现有安全设备难以有效保护网络：

- 无法检查网络内计算机的安全状况
- 缺乏对合法终端滥用网络资源的安全管理
- 无法防止恶意终端的蓄意破坏

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page4



大多数机构都使用身份管理及验证、授权和记帐(AAA) 机制来验证用户并为其分配网络访问权限，但对被验证用户的终端设备的安全状况几乎不起任何作用。如果不通过准确方法来评估设备‘状况’，即便是最值得信赖的用户也有可能在无意间通过受感染的设备或未得到适当保护的设备，将网络中所有用户暴露在巨大风险之中。

NAC使用网络基础设施对试图访问网络资源的所有设备执行安全策略检查，从而限制病毒、蠕虫和间谍软件等损害网络安全性。实施NAC的客户能够仅允许遵守安全策略的可信终端设备(PC、服务器及PDA等)访问网络，并控制不符合策略或不可管理的设备访问网络。

更多资料获取：http://learning.huawei.com/cn

NAC技术产生背景(续)

强化内防内控，从终端入手强化弱点管理：

- 终端接入控制：防止非法终端的接入，降低不安全终端的威胁
- 终端访问授权：防止合法终端越权访问，保护企业核心资源
- 终端安全健康性检查与策略管理：帮助企业落实安全管理制度
- 员工行为管理与违规审计：强化行为审计防止恶意终端破坏

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page5

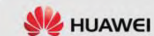


更多资料获取：http://learning.huawei.com/cn

NAC基本概念

- NAC (Network Access Control)：也称为网络接入控制，是思科最先提出的一种“端到端”的安全结构。
- 微软的NAP（网络访问保护）也是一种网络接入隔离控制技术，与NAC框架类似，服务器集成在Windows 2003 Server，客户端集成在Windows Vista客户端中。NAP服务器与客户端配合对于不符合当前系统运行状况要求的计算机进行强制受限网络访问。
- H3C的EAD (Endpoint Admission Defense)，称为端点准入防御。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page6



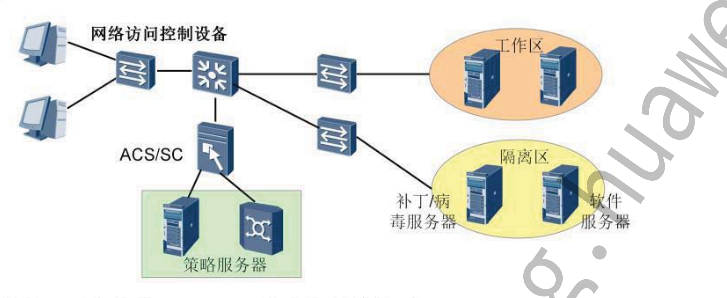
NAP还支持Windows 2008和 XPSP3。

- 网络访问保护NAP技术(Network Access Protection)是为微软下一代操作系统Windows Vista和Windows Server Longhorn设计的新的一套操作系统组件，它可以在访问私有网络时提供系统平台健康校验。NAP平台提供了一套完整性校验的方法来判断接入网络的客户端的健康状态，对不符合健康策略需求的客户端限制其网络访问权限。
- 为了校验访问网络的主机的健康，网络架构需要提供如下功能性领域:
  - 健康策略验证:判断计算机是否适应健康策略需求。
  - 网络访问限制:限制不适应策略的计算机访问。
  - 自动补救:为不适应策略的计算机提供必要的升级，使其适应健康策略。
  - 动态适应:自动升级适应策略的计算机以使其可以跟上健康策略的更新。

更多资料获取：http://learning.huawei.com/cn

NAC关键组件

NAC包含三个关键组件：通信代理、网络访问控制设备和策略服务器。



华为S系列交换机可用于网络访问控制设备。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page7



通信代理也就是安全客户端，是安装在用户终端系统上的软件，是对用户终端进行身份认证、安全状态评估以及安全策略实施的主体。

网络访问控制设备是企业网络中安全策略的实施点，起到强制用户准入认证、隔离不合格终端、为合法用户提供网络服务的作用。安全联动设备可以采用不同认证方式（如802.1x、MAC认证和Portal等）的端点准入控制。

策略服务器也就是管理服务器，NAC方案的核心是整合与联动，其中的安全策略服务器是NAC方案中的管理与控制中心，兼具用户管理、安全策略管理、安全状态评估、安全联动控制以及安全事件审计等功能。

华为公司的S系列交换机支持802.1X、Portal、MAC等认证控制技术，能够作为网络访问控制设备配合主流通信代理和策略服务器共同完成NAC控制，为企业网、园区网等提供安全可靠的访问控制。

更多资料获取：http://learning.huawei.com/cn

NAC认证方式

针对不同场景，NAC提供了灵活的接入控制方式：

- 802.1x认证接入（包括旁路认证）
- MAC地址认证接入
- Web认证接入

S9300除了提供上述NAC认证方式以外，还支持：

- 直接认证

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page8 

802.1X认证接入：

- IEEE 802.1x标准（以下简称802.1x）的主要内容是一种基于端口的网络接入控制（PortBased Network Access Control）协议。“基于端口的网络接入控制”是指在局域网接入控制设备的端口这一级对所接入的设备进行认证和控制。连接在端口上的用户设备如果能通过认证，就可以访问局域网中的资源；如果不能通过认证，则无法访问局域网中的资源。
- 802.1x协议仅关注接入端口的状态。当合法用户（根据帐号和密码）接入时，该端口打开；当非法用户接入或没有用户接入时，该端口处于关闭状态。认证的结果在于端口状态的改变，而不涉及通常认证技术必须考虑的IP地址协商和分配问题，是各种认证技术中最简化的实现方案。

MAC地址认证接入：

- MAC 地址认证是一种基于端口和MAC 地址对用户的网络访问权限进行控制的认证方法，它不需要用户安装任何客户端软件，用户名和密码都是用户设备的MAC 地址。网络接入设备在首次检测到用户的MAC 地址以后，即启动对该用户的认证。

WEB认证接入：

更多资料获取：http://learning.huawei.com/cn

- Web认证也称Portal认证，其基本原理是：用户首次打开浏览器，输入任何网址，都被强制重定向到Web服务器的认证页面，只有在认证通过后，用户才能访问网络资源。未认证用户只能访问特定的站点服务器。Web认证通过Web页面输入用户名和密码，使用Portal协议完成认证过程。
- Portal协议主要用于Web服务器和其他设备之间的信息交互。Portal协议基于客户端/服务器结构，采用UDP 作为传输协议。在Web认证中，Web认证服务器和S9300 之间的通信使用Portal协议，S9300为客户端。Web认证服务器从认证页面中将用户输入的用户名和密码提取后，通过Portal协议传送给S9300。

S9300单独提供的直接认证：

- 使能了接口的直接认证功能后，从该接口接入的用户将直接通过认证。
- 直接认证无全局开关，直接在接口下配置direct-authen使能，使能后不修改端口转发状态。
- 认证模块收到arp或者dhcp报文后，向服务器发送认证请求，该认证无需用户名密码，将直接进行授权。
- 直接认证支持动态下发VLAN和ACL。

- 本课程介绍802.1x认证方式，其它认证方式相对较简单，学员可自行参考相关手册。



更多资料获取：http://learning.huawei.com/cn



IEEE802 LAN/WAN委员会为解决无线局域网网络安全问题，提出了802.1x协议。后来，802.1x协议作为局域网端口的一个普通接入控制机制用在以太网中，主要解决以太网内认证和安全方面的问题。

更多资料获取：http://learning.huawei.com/cn

802.1x体系结构

802.1x系统为典型的Client/Server体系，包括三个实体：Supplicant system（客户端）、Authenticator system（设备端）和Authentication server system（认证服务器）

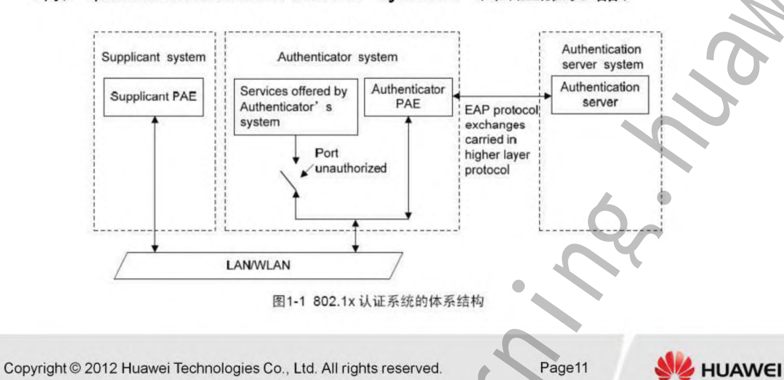


图1-1 802.1x 认证系统的体系结构

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page11 HUAWEI

客户端是位于局域网段一端的一个实体，由连接到该链接另一端的设备端对其进行认证。客户端一般为一个用户终端设备，用户通过启动客户端软件发起802.1x认证。客户端软件必须支持EAPOL（EAP over LANs，局域网上的EAP）协议。

设备端是位于局域网段一端的一个实体，用于对连接到该链接另一端的客户端进行认证。设备端通常为支持 802.1x协议的网络设备（如Quidway系列交换机），它为客户端提供接入局域网的端口，该端口可以是物理端口，也可以是逻辑端口。

认证服务器是为设备端提供认证服务的实体。认证服务器用于实现用户的认证、授权和计费，通常为 RADIUS服务器。该服务器可以存储用户的相关信息，例如用户的账号、密码以及用户所属的 VLAN、优先级、用户的访问控制列表等。

PAE（Port Access Entity，端口访问实体）：

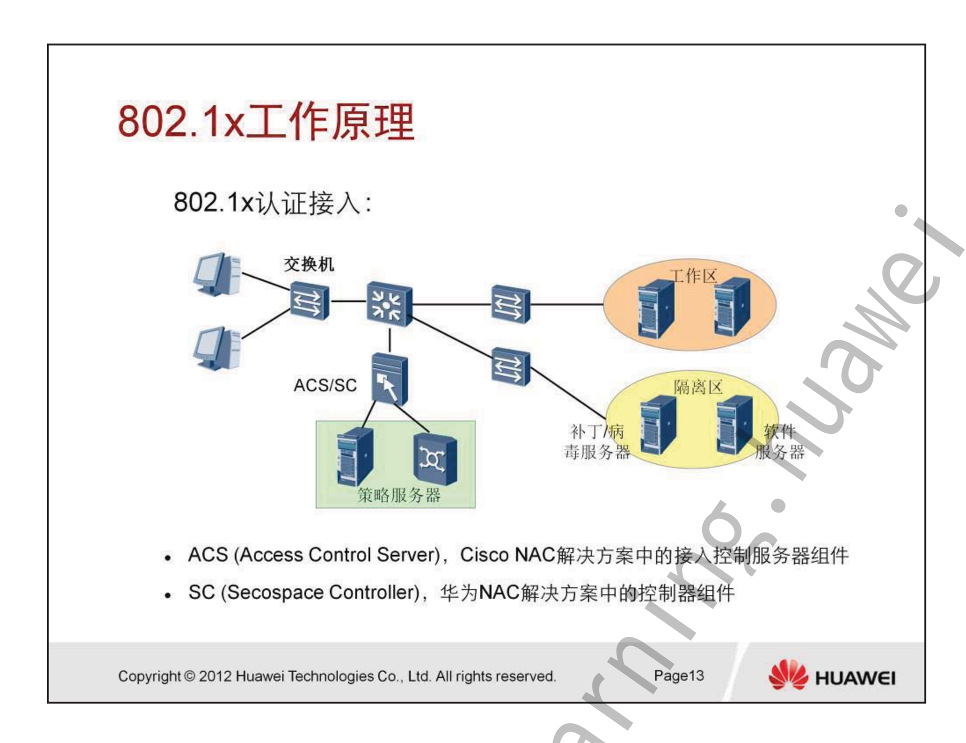
- PAE 是认证机制中负责执行算法和协议操作的实体。设备端 PAE 利用认证服务器对需要接入局域网的客户端执行认证，并根据认证结果相应地控制受控端口的授权/非授权状态。客户端PAE负责响应设备端的认证请求，向设备端提交用户的认证信息。客户端 PAE也可以主动向设备端发送认证请求和下线请求。

更多资料获取：http://learning.huawei.com/cn

受控端口：

- 设备端为客户端提供接入局域网的端口，这个端口被划分为两个虚拟端口：受控端口和非受控端口。
- 非受控端口始终处于双向连通状态，主要用来传递 EAPOL协议帧，保证客户端始终能够发出或接受认证。
- 受控端口在授权状态下处于连通状态，用于传递业务报文；在非授权状态下处于断开状态，禁止传递任何报文。
- 受控端口和非受控端口是同一端口的两个部分；任何到达该端口的帧，在受控端口与非受控端口上均可见。

更多资料获取：http://learning.huawei.com/cn



交换机检测到新的MAC上线后发起EAP认证请求。

- 若客户端没有响应：
  - 达到一定次数后认为未安装客户端软件，此时开放用户权限，仅允许用户访问隔离区。
  - 过一段时间（可配置）后再次发起探测。
- 若客户端已安装，则进行802.1x认证。认证通过后，PC和ACS之间建立HTTP链接，由ACS对PC机进行安全检查，检查通过后更新ACL，PC可访问工作区。

若客户端没有安装，但配置了MAC旁路认证功能，则将用户MAC作为用户名和密码进行认证。如果认证失败则使用户下线，并保持一段时间内不再发起认证和探测，超时后重新开始探测过程。此时如果用户下载了客户端，重新进行EAP认证，那么收到请求后，会先让该用户下线，再响应EAP的认证请求。

如果中间检测到PC感染了病毒，则下发ACL，使用户只能访问隔离区，并通过重定向URL要求用户更新病毒库或者打上相应补丁。

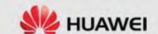
- 用户更新后，ACS检测到用户已经安全，则可通过RADIUS COA接口更新ACL，允许用户访问工作区。

更多资料获取：http://learning.huawei.com/cn

802.1x工作原理(续)

GUEST VLAN功能：

- 使能GUEST VLAN后，未获认证授权的用户，被临时加入GUEST VLAN，只能访问受限的部分资源（规划在GUEST VLAN中）。
- 未使能GUEST VLAN时，用户在通过认证之前，不能访问任何网络资源。
- 对于dot1x且端口配置为port-based情况，交换机将组播触发认证报文。
  - 若达到最大发送次数后仍无用户响应，端口就加入GUEST VLAN，端口下所有用户只能访问受限资源。
  - 若有其他用户认证成功，端口将退出GUEST VLAN，进入授权状态。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page14 

GUEST VLAN的使用需要规划好，尽量不和其他专用VLAN冲突，以保证用户在无访问权限时，只能访问病毒库/客户端下载/宣传页面/DHCP服务器等有限资源。

配置GUEST VLAN时请注意：

- 配置的Guest VLAN必须已经创建，且不能是接口的缺省VLAN。
- 接口下配置Guest VLAN以后，不能再配置将该接口加入该VLAN，也不能直接删除该VLAN。
- 不同的接口可以配置不同的Guest VLAN。
- 在同一视图下重复执行dot1x guest-vlan命令，新配置覆盖旧配置。

Guest VLAN的功能开启后：

- 交换机将在所有开启 802.1x功能的端口发送多播触发报文；
- 如果达到最大发送次数后，仍有端口尚未返回响应报文，则交换机将该端口加入到 Guest VLAN中；
- 之后属于该Guest VLAN中的用户访问该 Guest VLAN中的资源时，不需要进行 802.1x认证，但访问外部的资源时仍需要进行认证。

更多资料获取：http://learning.huawei.com/cn

802.1x工作原理(续)

静默功能：

- 为避免用户发送大量能启动认证流程的报文，导致设备不停地向RADIUS服务器发起认证请求，浪费设备和RADIUS服务器的处理资源，交换机需要提供静默功能。
- 启用静默功能后，在用户认证失败后的一段时间（静默周期，可配置）内，认证模块将不处理用户的认证报文。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page15



802.1x的静默功能在S9300上默认关闭，需通过命令开启。

为了避免静默功能使能后，用户认证失败1次就被静默，在S9300上可以配置静默前允许认证失败的次数大于1次。

S9300缺省情况下，802.1x用户在60秒内认证失败3次被静默。

在静默期间，S9300 丢弃该用户的802.1x认证请求。静默定时器的时长可以通过命令配置。

更多资料获取：http://learning.huawei.com/cn



EAP：Extensible Authentication Protocol，可扩展认证协议。

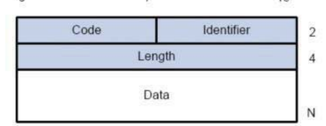
EAPOL：EAP over LANs，局域网上的EAP。

更多资料获取：http://learning.huawei.com/cn

EAP数据报

EAP数据报格式:

- 当EAPOL数据包Type域为EAP-Packet时，Packet Body为EAP数据包结构。



  - Code：指明EAP 包的4 种类型：**Request**、**Response**、**Success**、**Failure**
  - Identifier：辅助进行Request 和Response 消息的匹配
  - Length：EAP包的长度，包含Code、Identifier、Length和Data域，单位为字节
  - Data：EAP 包的内容，由Code 类型决定
    - Success和Fail类型的包没有Data域，相应Length域的值为4

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page17 

Code：代码字段，长度为一个字节。用于标识EAP的报文类型。

Identifier：标识字段，长度为一个字节，用于匹配请求报文和回应报文。

Length：长度字段，长度为两个字节，用于标识EAP报文的总长度，包括Code，Identifier，Length和Data字段。总长度最长为2的16次方为65535bit

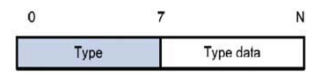
Data：数据字段，长度可变，根据代码字段的不同可以包含不同的内容。

更多资料获取：http://learning.huawei.com/cn

EAP数据报(续)

EAP请求(Request)和回应(Response)报文：

- 当EAP包为Request和Response类型时，其Data域的格式如下：



  - Type为EAP的认证类型
    - 值为1 时代表Identity，用来查询对方的身份；
    - 值为4 时代表MD5-Challenge，类似于PPP CHAP协议，包含质询消息
  - Type data的内容随Type值不同而不同

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page18



认证类型和认证信息就包含在类型字段（Type）和类型数据字段（Type-Data）中。

请求报文由验证者发给被验证端。

重传的请求报文必须使用相同的Identifier值，以区分其它新的请求报文。

被验证端必须使用回应报文答复请求报文。

回应报文只能用于答复一个收到的请求报文，而且不能被重传。

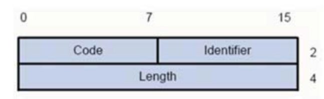
回应报文的Identifier值必须和所答复的请求报文的Identifier值相同。

更多资料获取：http://learning.huawei.com/cn

EAP数据报(续)

EAP成功(Success)和无效(Fail)报文:

- 当EAP包为Success和Fail类型时没有Data域，其Length域的值为4：



  - 成功报文由验证者发给被验证端，用于通告一个成功的认证结果。
  - 如果验证过程失败，验证者将向被验证端发送一个无效报文，通告一个无效的验证结果。
  - 成功报文和无效报文的Identifier值必须和回应报文的Identifier值一致。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page19



更多资料获取：http://learning.huawei.com/cn

EAPOL数据报

EAPOL报文的二层报文头：

DMAC	SMAC	TYPE	EAPOL	FCS

字段	内容
DMAC	01-80-C2-00-00-03
SMAC	端口的物理MAC地址
TYPE	PAE Ethernet Type，指示协议类型

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page20



当发送EAPOL数据帧的时候，目的地址为组MAC地址01-80-C2-00-00-03。

该组地址是IEEE802.1D保留的不能被交换机转发的组地址之一。

源MAC地址使用发送端口的物理MAC地址。

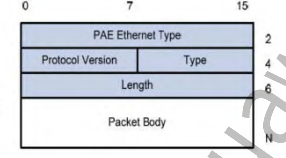
TYPE字段设置为88-8E，表示该数据帧中封装的是一个EAPOL数据帧。

更多资料获取：http://learning.huawei.com/cn

EAPOL数据报(续)

EAPOL数据报格式：

- PAE Ethernet Type：协议类型，值为0x888E。
- Protocol Version：指示EAPOL 帧的发送方所支持的协议版本号。
- Type：EAPOL数据包的类型。
- Length：表示数据长度，即 “Packet Body”字段的长度，单位为字节，0表示没有后面的数据域。
- Packet Body：表示数据内容，根据不同的Type有不同格式。



Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page21 HUAWEI

Version：版本字段，长度为一个字节，标识EAPOL数据帧发送者所使用的EAPOL协议版本号。目前版本号为“1”。

Type：类型字段，长度为一个字节，标识所传输的EAPOL数据帧的类型。

Length：所封装的EAP数据包的长度。0标识没有封装EAP数据包。

Packet Body：EAP数据包被封装在此字内。

更多资料获取：http://learning.huawei.com/cn

EAPOL数据报(续)

EAPOL数据包类型:

Type值	报文类型
0x00	EAP报文（EAP-Packet），用于承载认证信息
0x01	EAPOL开始报文（EAPOL-Start），发起认证
0x02	EAPOL注销报文（EAPOL-Logoff），退出请求
0x03	EAPOL信息报文（EAPOL-Key）
0x04	EAPOL告警报文（EAPOL-Encapsulated-ASF-Alert）

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.
Page22


EAP报文：该EAPOL数据帧包含一个EAP报文。

EAPOL开始报文：由请求者发送给验证者，触发验证者启动认证协议。

EAPOL注销报文：由请求者发送给验证者，标识一个显式的注销请求。

EAPOL信息报文：验证者和请求者的802.1X可选能力交换信息所使用。

EAPOL告警报文：用于传输告警标准论坛（ASF）定义的告警信息，例如SNMP Trap信息。

http://learning.huawei.com/cn

更多资料获取：

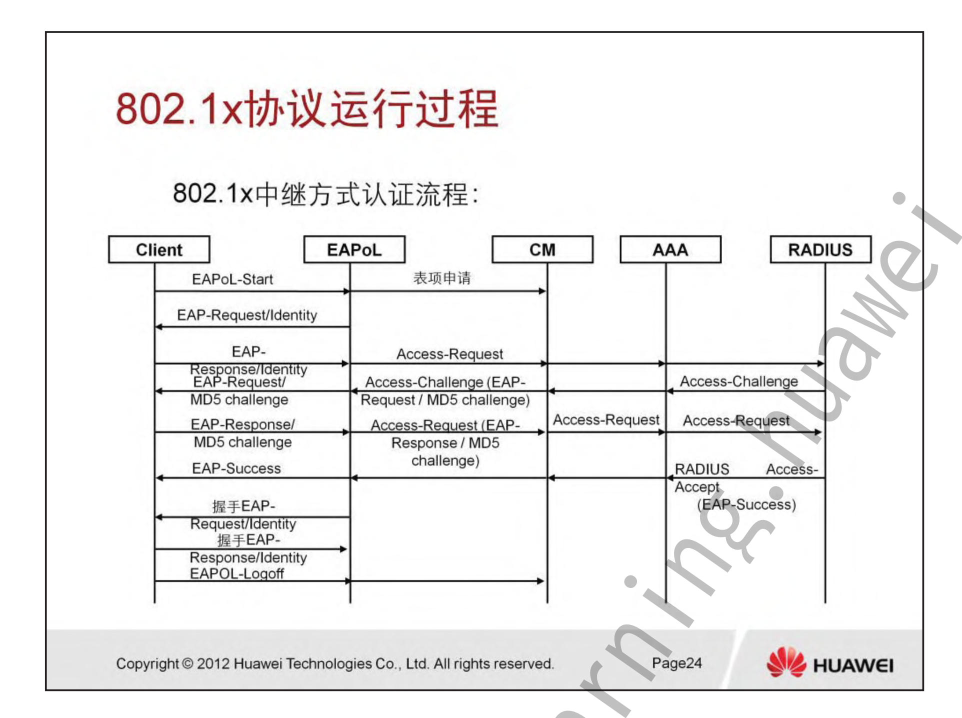


目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page23



更多资料获取：http://learning.huawei.com/cn



802.1x的认证流程包括中继方式和终结方式两种。

- 中继方式的认证中，交换机将EAP请求回应和MD5 challenge请求进行透传，所以我们称这种方式为中继方式，也称为透传认证方式。

EAP 中继方式认证过程如下（设备指的是网络接入设备）：

- 1. EAP 客户端发送EAP-Start报文给设备。
- 2. 设备发送EAP-Request/Identity报文给客户端。
- 3. 客户端回应EAP-Response/Identity报文，设备透传报文给RADIUS服务器。
- 4. 设备收到RADIUS挑战报文后，发送EAP挑战报文EAP-Request/MD5-Challenge给客户端。
- 5. 客户端回应EAP-Response/MD5-Challenge报文，设备透传报文给RADIUS服务器。
- 6. 设备认证成功后，通知客户端认证成功，端口打开。
- 7. 客户端在线过程中，设备通过EAP握手报文进行探测客户端是否保持在线。

更多资料获取：http://learning.huawei.com/cn

802.1x协议运行过程(续)

802.1x终结方式认证流程:

Client　EAPoL　CM　AAA　RADIUS

- Client → EAPoL: EAPoL-Start
- EAPoL → CM: 表项申请
- EAPoL → Client: EAP-Request/Identity
- CM → EAPoL: 申请成功
- Client → EAPoL: EAP-Response/Identity
- EAPoL → Client: EAP-Request / MD5 challenge
- Client → EAPoL: EAP-Response / MD5 challenge
- EAPoL → CM: Access-Request (EAP-Response/MD5 challenge)
- CM → AAA: Access-Request
- AAA → RADIUS: Access-Request
- RADIUS → AAA: RADIUS Access-Accept (EAP-Success)
- EAPoL → Client: EAP-Success
- EAPoL → Client: 握手EAP-Request/Identity
- Client → EAPoL: 握手EAP-Response/Identity
- Client → EAPoL: EAPOL-Logoff

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page25 HUAWEI

EAP 终结认证过程如下（设备指的是网络接入设备）：

- 1. EAP客户端发送EAP-Start报文给设备。
- 2. 设备发送EAP-Request/Identity报文给客户端。
- 3. 客户端回应EAP-Response/Identity报文，报文携带用户名信息。
- 4. 设备发送EAP-Request/MD5-Challenge报文给客户端。
- 5. 客户端回应EAP-Response/MD5-Challenge报文，设备获取客户端的密码信息。
- 6. 设备携带用户账户信息，到AAA系统进行认证。
- 7. 认证通过后设备通知客户端认证成功，端口打开。
- 8. 设备通过EAP探测判断EAP客户端是否维持在线。

更多资料获取 http://learning.huawei.com/cn



目 录

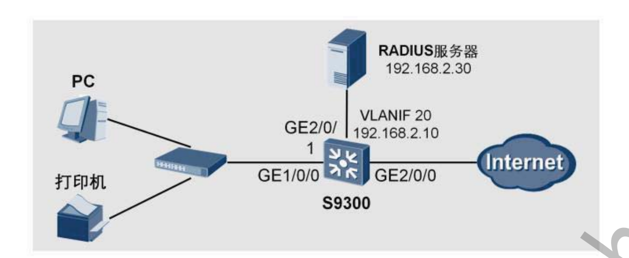
Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page26

HUAWEI

更多资料获取：http://learning.huawei.com/cn

802.1x业务配置基本组网



需要注意，NAC特性部署时与周边交互的模块比较多，各模块间协同工作需要正确配置，若配置不当，常会导致认证失败。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page27 

业务要求：

- 对GE1/0/0接口上的接入用户进行802.1x认证，以控制其访问Internet，接入控制方式采用缺省方式，即基于MAC的接入控制方式。
- 使用RADIUS服务器完成认证。
- GE1/0/0接口最大接入用户数为100。
- 对GE1/0/0接口下的打印机采用MAC旁路认证。

对于某些特殊终端，例如打印机等，无法使用和安装802.1x终端软件，可以通过基于MAC的旁路认证方式进行认证。

MAC的旁路认证方式指当终端进行802.1x认证失败后，把它的MAC地址作为用户名和密码上送RADIUS服务器进行认证。

更多资料获取 http://learning.huawei.com/cn

802.1x配置准备

配置思路：

- 配置RADIUS服务器模板
- 配置AAA认证模板
- 配置域
- 配置802.1x认证

需要准备如下数据：

- RADIUS服务器IP地址、认证端口号
- RADIUS服务器密钥为hello，重传次数为2
- AAA认证方案web1
- RADIUS服务器模版rd1
- 域isp1

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page28



配置前需要预先确定802.1x业务的配置思路和准备数据，然后才能进行后续配置。

更多资料获取：http://learning.huawei.com/cn

802.1x业务配置

配置步骤：

- 配置RADIUS服务器模板
  - [Quidway] radius-server template rd1
  - [Quidway-radius-rd1] radius-server authentication 192.168.2.30 1812

    #配置RADIUS主用认证服务器的IP地址、端口
  - [Quidway-radius-rd1] radius-server shared-key hello

    #配置RADIUS服务器密钥
  - [Quidway-radius-rd1] radius-server retransmit 2

    #配置重传次数
  - [Quidway-radius-rd1] quit

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page29 

注意事项：

- 若仅使用本地认证，不需要关注RADIUS配置
- 交换机是RADIUS的客户端，以上仅为客户端配置。要使配置真正生效，还需要对服务器端进行配置
- 通信密钥需要与服务器协商好，否则通信不能完成
- 如果和服务器中的的域名不一致，可以指示系统从用户名中去除用户域名后再将之传给RADIUS服务器。命令如下：
  - [Quidway-radius-rd1] user-name-format without-domain

更多资料获取：http://learning.huawei.com/cn

802.1x业务配置(续)

配置认证方案，认证方案web1，认证方法为RADIUS

- [Quidway] aaa
- [Quidway-aaa] authentication-scheme web1
- [Quidway-aaa-authen-1] authentication-mode radius
- [Quidway-aaa-authen-1] quit

配置isp1域，绑定认证方式和RADIUS服务器模板

- [Quidway-aaa] domain isp
- [Quidway-aaa-domain-isp1] authentication-scheme web1
- [Quidway-aaa-domain-isp1] radius-server rd1

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page30



域绑定认证方案、认证模板：

- 认证方案指明认证类型为RADIUS。
- 通过认证模板查询到相应的RADIUS。

更多资料获取：http://learning.huawei.com/cn

802.1x业务配置(续)

配置802.1x认证

- [Quidway] dot1x

  #全局使能802.1x认证

- [Quidway] interface gigabitethernet1/0/0
- [Quidway-GigabitEthernet1/0/0] dot1x

  #在接口下使能802.1x认证

- [Quidway-GigabitEthernet1/0/0] dot1x max-user 100

  #配置允许接入的最大用户数

- [Quidway-GigabitEthernet1/0/0] dot1x mac-bypass

  #配置MAC旁路认证

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.



缺省情况下，全局未使能802.1x认证功能。

如果当前接口下有802.1x在线用户，此接口下不允许去使能802.1x认证功能。

配置允许接入的最大用户数：

- 根据S9300设备类型不同，整机允许接入的NAC最大用户数也不同。规格为：8192*接口板槽位数。
- 若接口下配置dot1x port-method命令指定认证方式为基于接口（port）对接入用户进行认证，则接口的最大接入用户数变为1。此时不能执行dot1x max-user命令，需要undo dot1x port-method才能再进行最大用户数的配置。
- 当接口下的在线用户数大于待配置的最大用户数时，该配置会使接口下所有用户下线。系统会给出警告信息，提醒您是否继续：“Warning: The total number of online users is greater than the limit, so all the online users will go offline. Are you sure to continue?[Y/N]：”。
- 在同一视图下重复执行dot1x max-user命令，新配置会覆盖旧配置。

更多资料获取：http://learning.huawei.com/cn

802.1x业务配置(续)

其它可选配置：

- dot1x authentication-method {chap | eap | pap }

  #配置dot1x认证模式
- dot1x dhcp-trigger

  #使能DHCP报文触发dot1x认证功能
- dot1x handshake

  #开启定时握手功能（默认使能）
- dot1x quiet-period

  #开启dot1x静默功能
- dot1x retry

  #配置报文交互时交换机向客户端重发报文的最大次数

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.





在实际业务配置过程中，需要根据客户需求，选择具体可选配置项。因此需要了解这些可选项的默认值。

配置dot1x认证模式dot1x authentication-method {chap | eap | pap }

- PAP（Password Authentication Protocol）是一种两次握手认证协议，它采用明文方式传送口令；
- chap（Challenge Handshake Authentication Protocol）为终结方式（默认），是一种三次握手认证协议，它只在网络上传输用户名，而并不传输口令。相比之下，CHAP认证保密性较好，更为安全可靠。
- eap为中继方式，交换机直接把802.1x用户的认证信息以EAP报文发送给RADIUS服务器完成认证，而无须将EAP报文转换成标准的RADIUS报文后再发给RADIUS服务器来完成认证。

使能DHCP报文触发dot1x认证功能 dot1x dhcp-trigger

- 缺省情况下，DHCP 不触发接入用户的身份认证。
- 在执行此命令后，若用户未通过认证，则无法从DHCP Server获得动态IP 地址。

开启dot1x静默功能dot1x quiet-period

- 缺省情况下，802.1x用户在60秒内认证失败3次被静默。

更多资料获取:http://learning.huawei.com/cn

802.1x业务配置(续)

其它可选配置(续):

- dot1x timer
  #配置各类定时器，确保有序交互
- dot1x guest-vlan
  #配置接口guest vlan功能
- dot1x port-control {auto | authorized-force | unauthorized-force }
  #配置端口控制模式
- dot1x port-method { mac | port }
  #配置端口接入方式，即基于MAC（用户）还是基于端口
- dot1x reauthenticate
  #配置重认证，将该端口下通过认证的用户定期进行重认证

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page33 

dot1x port-control { auto | authorized-force | unauthorized-force } 命令用于配置端口控制模式:

- 默认auto，自动识别模式（auto）：接口初始状态为非授权状态，仅允许收发EAPoL报文，不允许用户访问网络资源；如果认证通过，则接口切换到授权状态，允许用户访问网络资源。
- authorized-force：强制授权模式（authorized-force）：接口始终处于授权状态，允许用户不经认证授权即可访问网络资源。
- 强制非授权模式（unauthorized-force）：接口始终处于非授权状态，不允许用户访问网络资源。

dot1x reauthenticate命令可以对用户进行定期重认证，无需用户介入。

dot1x guest-vlan:

- 当Guest VLAN 功能开启后，S9300 向所有开启802.1x功能的端口广播主动认证报文，如果达到最大重认证次数后，仍有端口上未返回响应报文，则S9300 将该端口加入到Guest VLAN 中。之后属于该Guest VLAN 中的用户访问该Guest VLAN 中的资源时，不需要进行802.1x认证，但访问外部的资源时仍需要进行认证。从而满足了允许未认证用户访问某些资源的需求。

更多资料获取：http://learning.huawei.com/cn

- 配置的Guest VLAN 不能是接口的缺省 VLAN。
- 缺省情况下，接口下未配置Guest VLAN。

dot1x port-method { mac | port }:

- 缺省情况下，接口的接入模式为基于MAC 地址。
- 当有802.1x用户在线时，不允许执行本命令更改接口接入控制方式。

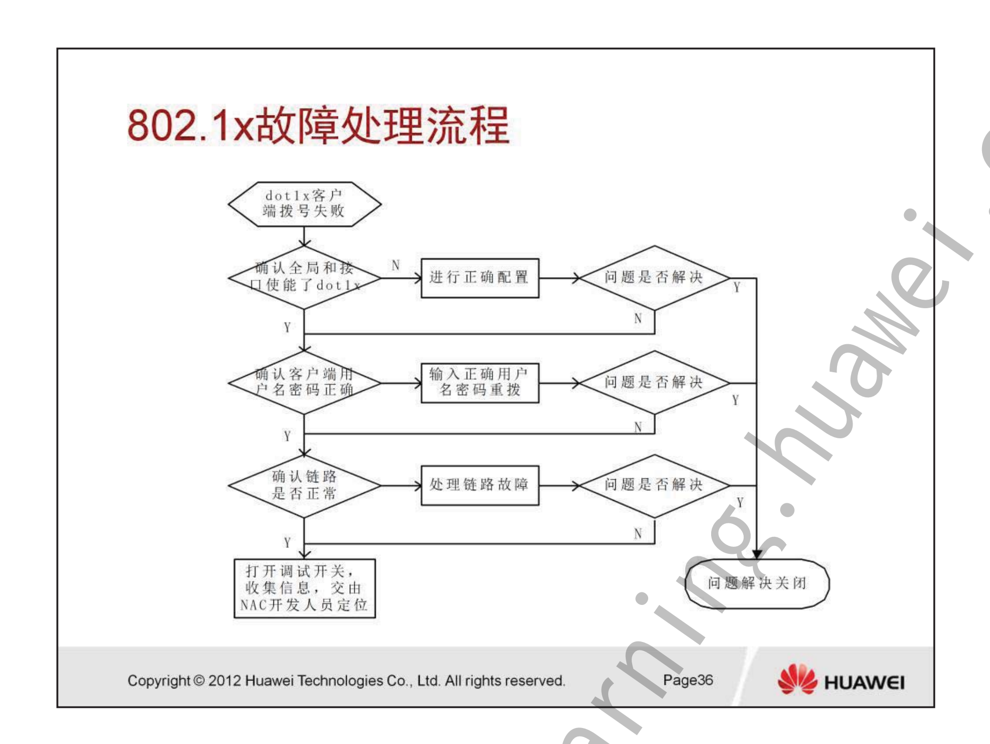
更多资料获取：http://learning.huawei.com/cn



目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page35 HUAWEI

更多资料获取：http://learning.huawei.com/cn



在此故障流程中默认radius正常工作。

如果radius不正常工作，则需要查看是否能够ping通radius serve，radius是否配置正确。如果还是不能解决问题，需要打开debugging radius进行信息采集，发给华为专业人员定位。

更多资料获取：http://learning.huawei.com/cn

802.1x基本故障诊断

查看dot1x是否使能

- display dot1x
  - 若显示Global 802.1x is Disabled，则需要在system视图下输入命令dot1x以开启dot1x特性全局开关。
- display dot1x interface GigabitEthernet 2/0/2
  - 若显示802.1x protocol is Disabled，则需在接口下使能dot1x功能。在system视图下输入命令：
    - [system]interface GigabitEthernet 2/0/2
    - [system-GigabitEthernet2/0/2]dot1x

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page37



在进行802.1x故障诊断时，可以根据这样的步骤来逐步排查故障点。

也可以直接查看配置，如果当前配置无dot1x，则增加配置。

更多资料获取：http://learning.huawei.com/cn

802.1x基本故障诊断(续)

确认用户名密码是否正确

- 确认用户名和密码在RADIUS服务器的合法帐号列表
- 否则使用正确的用户名密码重新拔号

确认链路是否正常

- display interface GigabitEthernet 2/0/2
- 确认接口UP，且有流量
- 否则检查线路是否有故障，VLAN配置是否正确，以确保客户端拔号后，端口有报文收发。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page38 

在进行802.1x故障诊断时，可以根据这样的步骤来简单排查故障点。

当端口的物理状态为down时，则数据链路层存在故障。

- 如果是光口，测试收发光功率，确定问题。
- 如果是RJ45，则使用测线器测试，确定问题。
- 如果线路没有问题，再查看两端的端口工作模式是否一致，并修改为一致。建议关闭自协商。

更多资料获取：http://learning.huawei.com/cn



802.1x和NAC的关系?

EAP报文有哪四种类型?

802.1x的接入认证有哪两种方式?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page39 

802.1x和NAC的关系?

- 802.1x是NAC接入认证方式中的一种。NAC接入认证还包括MAC地址认证、WEB认证和直接认证。

EAP报文有哪四种类型?

- 包括请求、回应、成功和失效四种报文，通过Code字段识别。

802.1x的接入认证在交换机上有哪两种方式?

- 有中继方式和终结方式两种认证。

更多资料获取：http://learning.huawei.com/cn



谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



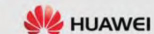
前 言

随着网络规模的扩大和网络复杂度的提高，网络配置越来越复杂，计算机位置变化（如便携机或无线网络）和计算机数量超过可分配的IP地址，造成IP地址变化频繁以及IP地址不足的情况。

DHCP是Dynamic Host Configuration Protocol的简称，又称为动态主机配置协议，是一种对用户进行集中的动态管理和配置的技术。

DHCP技术保证了IP地址的合理分配问题，从而避免了IP地址的浪费，提高了整网的IP地址使用率。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1



在常见的小型网络中（例如家庭网络和学生宿舍网），网络管理员都是采用手工分配IP地址的方法。而到了中、大型网络，如企业网，采用手工配置IP地址，需要详细的规划，工作量大，且不好管理，用户擅自修改地址，造成各种故障，带来安全隐患，尤其用户把地址修改为关键的服务器或者网关地址，将会导致严重的后果。用户不了解网络情况，没有专业的知识，配置IP有一定难度。

使用DHCP避免用户手工配置IP地址造成的地址冲突。降低对用户的要求。

DHCP服务优点：网络管理员可以验证IP地址和其它配置参数，而不用去检查每个主机；DHCP不会同时租借相同的IP地址给两台主机；DHCP管理员可以约束特定的计算机使用特定的IP地址；可以为每个DHCP作用域设置很多选项；客户机在不同子网间移动时不需要重新设置IP地址。

DHCP服务缺点：DHCP不能发现网络上非DHCP客户机已经在使用的IP地址；当网络上存在多个DHCP服务器时，一个DHCP服务器不能查出已被其它服务器租出去的IP地址；DHCP服务器不能跨路由器与客户机通信，除非路由器允许BOOTP转发。

更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 掌握DHCP的基本原理
- 掌握DHCP Server
- 掌握DHCP Relay
- 掌握DHCP Snooping
- 掌握DHCP业务配置

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page2

HUAWEI

更多资料获取：http://learning.huawei.com/cn



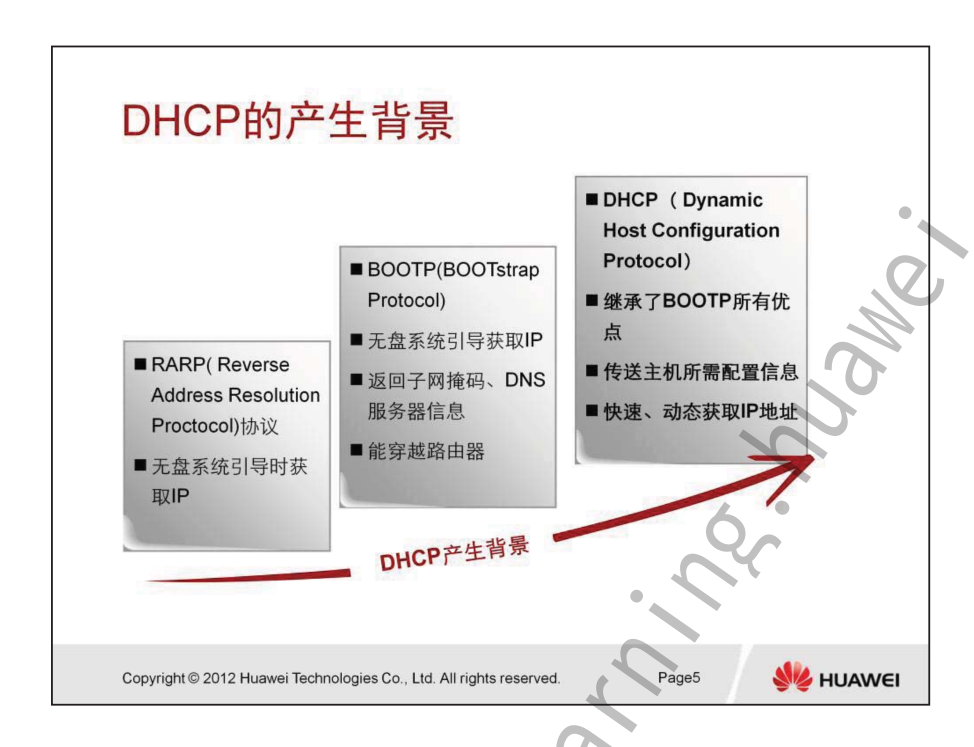
更多资料获取：http://learning.huawei.com/cn



DHCP的基本原理

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page4 HUAWEI

更多资料获取：http://learning.huawei.com/cn



DHCP的由来：RARP→BOOTP →DHCP

RARP(Reversed Address Resolution Protocol)协议：

- RARP是许多无盘系统在引导时获取IP地址的。 RARP请求在网络上广播，在报文中标识发送端的硬件地址，以请求相应的IP地址，应答通常是单播方式的。RARP的分组格式基本上与ARP相同。

BOOTP（BOOTstrap Protocol）协议：

- BOOTP使用UDP报文封装,也是无盘系统用来获取IP地址的方法。它还能返回其他信息，如路由器的IP地址，客户的子网掩码和DNS服务器地址等。与RARP相比，它除了能够获得更多的信息，还可以穿越路由器。

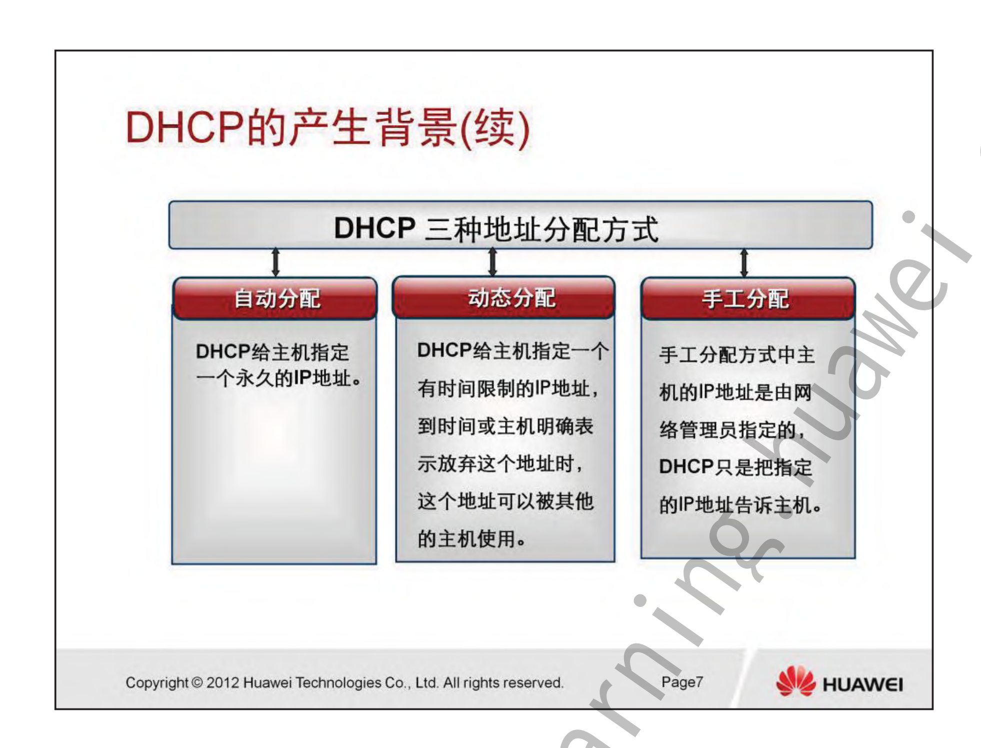
DHCP（Dynamic Host Configuration Protocol）协议:

- DHCP是从BOOTP的基础上发展过来的，他继承了BOOTP的许多优点，最主要的改进是对用户IP地址的动态分配。它的适用范围不光是无盘系统，更多的用在不同的网络间移动的系统上。
- DHCP从两个方式上扩充了BOOTP，第一，DHCP可使计算机用一个消息获取它所需要的所有配置信息，即传送配置信息的协议；第二，DHCP允许计算机快速、动态的获取IP地址，

更多资料获取：http://learning.huawei.com/cn

- 即动态分配IP地址的机制。
- DHCP建立在client-server模型上。其中指定的DHCP server分配网络地址并向动态配置的主机传送配置参数。只有当系统管理员明确的配置主机作为DHCP服务器时，主机才能作为服务器来工作。

更多资料获取：http://learning.huawei.com/cn



DHCP服务器支持三种类型的地址分配：

- 自动分配：在自动分配中，不需要进行任何的IP地址手工分配。当DHCP客户机第一次向DHCP服务器租用到IP地址后，这个地址就永久地分配给了该DHCP客户机，而不会再分配给其他客户机。
- 动态分配：当DHCP客户机向DHCP服务器租用IP地址时，DHCP服务器只是暂时分配给客户机一个IP地址。只要租约到期，这个地址就会还给DHCP服务器，以供其他客户机使用。如果DHCP客户机仍需要一个IP地址来完成工作，则可以再要求另外一个IP地址。
- 手动分配：在手动分配中，网络管理员在DHCP服务器上通过手工方法配置DHCP客户机的IP地址。当DHCP客户机要求网络服务时，DHCP服务器把手工配置的IP地址传递给DHCP客户机。

在这三种方式中，只有动态分配的方式可以对已经分配给主机但现在此主机已经不用的IP地址重新加以利用。这样，在给一台临时连入网络的主机分配地址或者在一组不需要永久的IP地址的主机中共享一组有限的IP地址时，动态分配就会显得特别有用。当一台新主机要永久的接入一个网络时，而网络的IP地址非常有限，为了将来这台主机被淘汰时能回

更多资料获取：http://learning.huawei.com/cn

收IP地址，这种情况下动态分配也是一个很好的选择。

IP地址分配的优先次序

- DHCP服务器按照如下次序为客户端选择IP地址。
- DHCP服务器的数据库中与客户端MAC地址静态绑定的IP地址；
- 客户端以前曾经使用过的IP地址，即客户端发送的DHCP_DISCOVER报文中请求IP地址选项（Requested IP Addr Option）的地址；
- 在DHCP地址池中，顺序查找可供分配的IP地址，最先找到的IP地址；
- 如果在DHCP地址池中未找到可供分配的IP地址，则依次查询超过租期、发生冲突的IP地址，如果找到可用的IP地址，则进行分配，否则报告错误。

更多资料获取：http://learning.huawei.com/cn

DHCP的产生背景(续)

DHCP与BOOTP协议比较:

相同点	不同点
■ Client/Server 模式	■ BOOTP运行在静态环境中
■ 客户端提出配置申请	■ 主机需配置BOOTP参数文件
■ 服务器端响应配置请求	■ 文件长时间不变
■ 采用UDP封装	■ DHCP允许主机快速、动态获取IP地址
■ 相同的报文格式	

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page9



DHCP协议与BOOTP协议的比较:

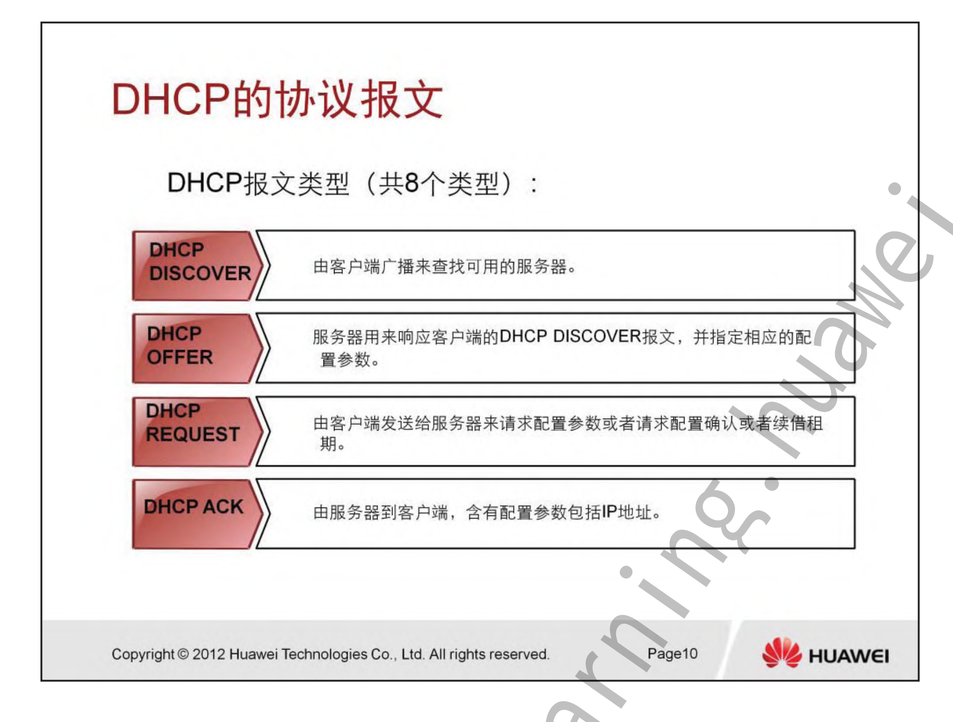
- 相同点：DHCP和BOOTP都采用客户/服务器通信模式，由客户端向服务器提出配置申请（包括分配的IP地址、子网掩码、缺省网关等参数），服务器根据策略返回相应配置信息，两种报文都采用UDP进行封装，并使用基本相同的报文结构。DHCP和BOOTP均使用两个知名端口：服务器为67，客户端为68。
- 不同点：BOOTP运行在相对静态（每台主机都有固定的网络连接）的环境中，管理员为每台主机配置专门的BOOTP参数文件，该文件会在相当长的时间内保持不变。DHCP允许计算机快速、动态地获取IP地址，而不是静态为每台主机指定地址。

DHCP从两方面对BOOTP进行了扩展:

- DHCP可使计算机仅用一个消息就获取它所需要的所有配置信息。
- DHCP允许计算机快速、动态地获取IP地址，而不是静态为每台主机指定地址。



更多资料获取：http://learning.huawei.com/cn

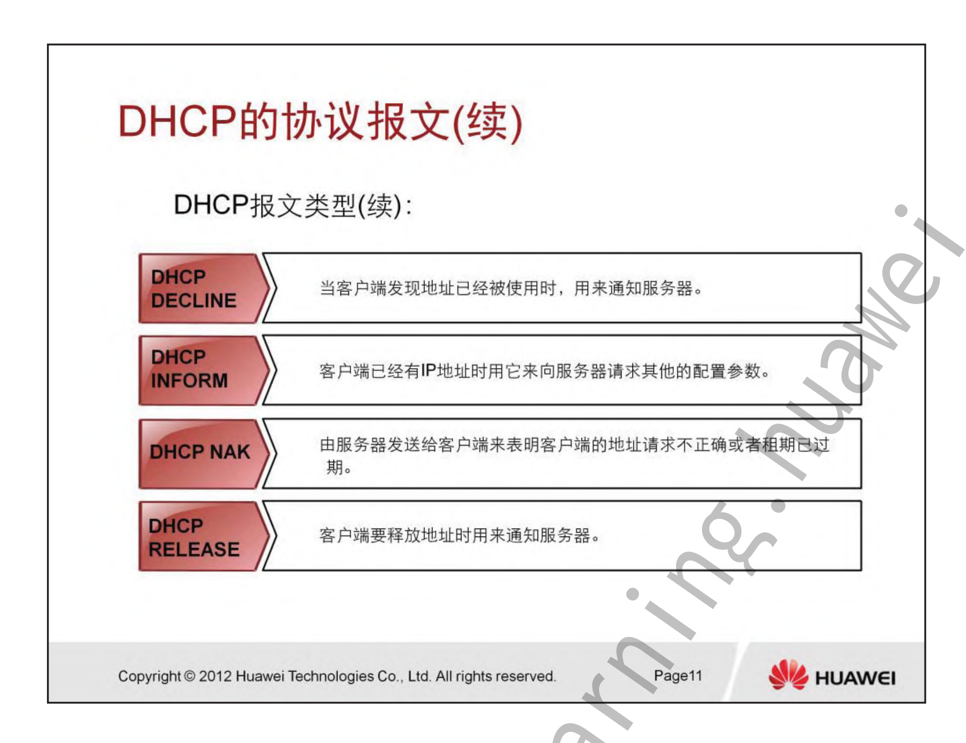


任何动态的协议都有自己一套规定的语言即协议，DHCP也不例外。

DHCP报文共有8个类型，具体如下：

- DHCP DISCOVER：这是DHCP客户端首次登录网络时进行DHCP过程的第一个报文，用来寻找DHCP服务器。
- DHCP OFFER：服务器用来响应客户端的DHCPDISCOVER报文，并指定相应的配置参数。
- DHCP REQUEST：由客户端发送给服务器来请求配置参数或者请求配置确认或者续借租期。此报文用于以下三种用途。
    - 客户端初始化后，发送广播的DHCPREQUEST报文来回应服务器的DHCPOFFER报文。
    - 客户端重启初始化后，发送广播的DHCPREQUEST报文来确认先前被分配的IP地址等配置信息。
    - 当客户端已经和某个IP地址绑定后，发送单播的DHCPREQUEST报文来延长IP地址的租期。
- DHCP ACK：服务器对客户端的DHCPREQUEST报文的确认响应报文，客户端收到此报文后，才真正获得了IP地址和相关的配置信息。

更多资料获取：http://learning.huawei.com/cn



DHCP报文的另外4个类型：

- DHCP DELINE：当客户端发现服务器分配给它的IP地址发生冲突时会通过发送此报文来通知服务器，并且会重新向服务器申请地址。
- DHCP NAK：服务器对客户端的DHCP REQUEST报文的拒绝响应报文，比如服务器对客户端分配的IP地址已超过使用租借期限或者客户端移到了另一个新的网络。
- DHCP INFORM：客户端已经获得了IP地址，发送此报文的目的是为了从服务器获得其他的一些网络配置信息，比如网关地址、DNS服务器地址等。
- DHCP RELEASE：客户端可通过发送此报文主动释放服务器分配给它的IP地址，当服务器收到此报文后，可将这个IP地址分配给其它的客户端。

更多资料获取：http://learning.huawei.com/cn

DHCP的协议报文(续)

DHCP报文格式

OP (1)	Htype (1)	Hlen (1)	Hops (1)
Xid (4)			
Secs (2)		Flags (2)	
Client IP address (4)			
Your IP address (4)			
Server IP address (4)			
Gateway IP address (4)			
Client Hardware Address (16)			
Server Name (64)			
File (128)			
Options (可变)			

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page12



OP:操作码 (1=bootrequest ,2=bootreply)

Htype: 硬件地址类型 (1=10mb ethernet)

Hlen: 硬件地址长度 (ethernet 为10)

Hops:表示当前的DHCP报文经过的DHCP Relay的数目。该字段由客户端设置为0，每经过一个DHCP Relay时，该字段加1。此字段的作用是限制DHCP报文所经过的DHCP中继数目。服务器和客户端之间的DHCP中继不能超过4次，也就是Hops值不能大于4，否则DHCP报文将被丢弃。

Xid: 传输ID,在同 服务器的交互中,由客户机所选择

Secs: 客户机所使用的地址,在最近一次地址获取/地址更新后所经过的时间

Flags:此字段在BOOTP中保留未用，在DHCP中表示标志字段。只有标志字段的最高位才有意义，其余的位均被置为0。最左边的字段被解释为广播响应标志位，内容如下所示:

- 0：客户端请求服务器以单播形式发送响应报文
- 1：客户端请求服务器以广播形式发送响应报文

Client IP address: 客户机在BOUND,RENEW或REBINDING状态所使用,可以用来回应ARP请求报文

更多资料获取：http://learning.huawei.com/cn

Client IP Address： 该字段表示客户端的IP地址。可以是服务器分配给客户端的IP地址或者客户端已有的IP地址。客户端在初始化状态时没有IP地址，此字段为0.0.0.0。IP地址0.0.0.0仅在采用DHCP方式的系统启动时允许本主机利用它进行临时的通信，并且永远不是有效目的地址。

Your IP address: 服务器给客户机分配的IP地址。

Server IP address:该字段表示服务器IP地址。

Gateway IP address:该字段表示第一个DHCP中继的IP地址。当客户端发出DHCP请求时，如果服务器和客户端不在同一个网络中，那么第一个DHCP中继在转发这个DHCP请求报文时会把自己的IP地址填入此字段。服务器会根据此字段来判断出网段地址，从而选择为用户分配地址的地址池。服务器还会根据此地址将响应报文发送给此DHCP中继，再由DHCP中继将此报文转发给客户端。若在到达DHCP服务器前经过了不止一个DHCP中继，那么第一个DHCP中继后的中继不会改变此字段，只是把Hops的数目加1。

Client hardware address: 该字段表示客户端的MAC地址，此字段与前面的“Hardware Type”和“Hardware Length”保持一致。当客户端发出DHCP请求时，将自己的硬件地址填入此字段。对于以太网，当“Hardware Type”和“Hardware Length”分别为“1”和“6”时，此字段必须填入6字节的以太网MAC地址。

Server name:该字段表示客户端获取配置信息的服务器名字。此字段由DHCP Server填写，是可选的。如果填写，必须是一个以0结尾的字符串。 缺省为空。

File: 启动文件的名字,在DHCPOFFER报文中给出全名。

Options:该字段表示DHCP的选项字段，至少为312字节。DHCP通过此字段包含了服务器分配给终端的配置信息，如网关IP地址，DNS服务器的IP地址，客户端可以使用IP地址的有效租期等信息。

更多资料获取：http://learning.huawei.com/cn

DHCP的协议报文(续)

- DHCP报文中的option字段，采用“CLV”方式构成。
  - Code:标识号，唯一标识后面的信息内容，占1byte
  - Length:长度，表示后面信息内容的长度，占1byte
  - Value:信息内容，其长度为length所指定，以byte为单位

0	8	16
代码	长度	数据

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page14 HUAWEI

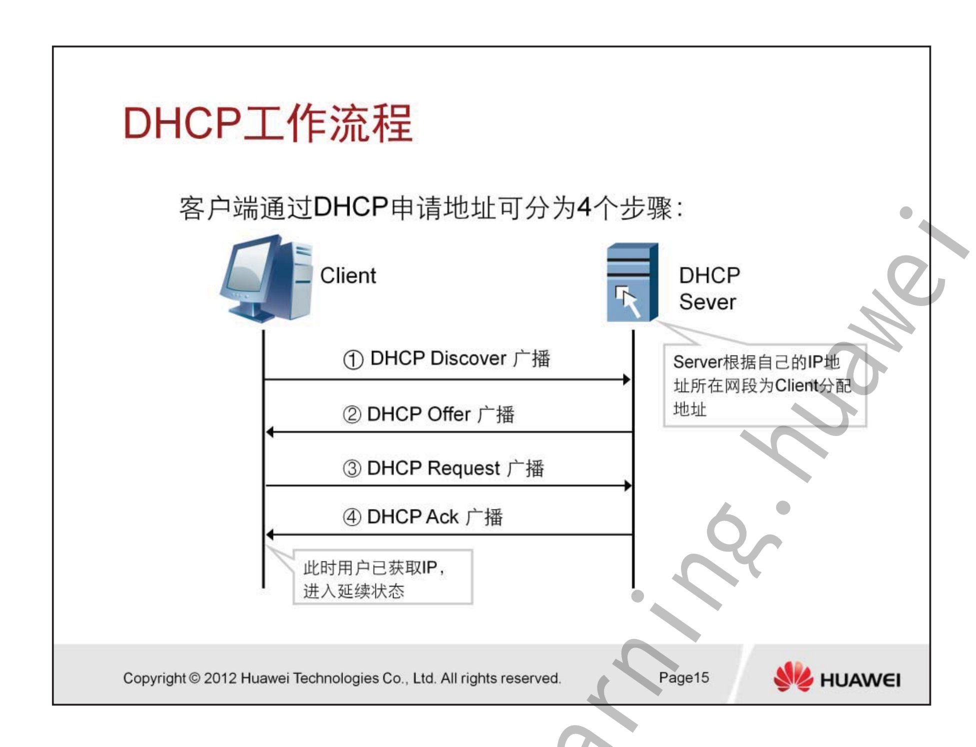
DHCP报文中的option字段，采用“CLV”方式构成，即Code:标识号，唯一标识后面的信息内容，占1byte；Length:长度，表示后面信息内容的长度，占1byte；Value:信息内容，其长度为length所指定，以byte为单位。这种方式非常灵活，当需要新的信息时，可以按照这种编码方式申请新的option即可。此方式具有可扩展性，在协议上有广泛的应用。

常见的option类型有：

- 报文类型: C=53, L=1, V = 1--8，表示DHCP报文类型
- Router IP: C=3 , L=IP地址长度, V = client的默认网关的IP地址
- DNS IP : C=6 , L=IP地址长度的倍数, V = client的DNS服务器的IP地址序列
- WINS IP : C=44, L=IP地址长度的倍数, V = client的WINS服务器的IP地址序列

DHCP提供了在TCP/IP网络上传输配置参数的框架，在DHCP客户端和服务器端可以用选项代码传送双方约定的配置参数和控制信息。

更多资料获取：http://learning.huawei.com/cn



DHCP工作流程：

- 客户端发送DHCP Discover 广播报文即目的地址为255.255.255.255，在网络上寻找DHCP Server。
- 网络的的DHCP服务器响应客户端的请求，可能有多个DHCP Sever 响应，以广播的方式进行。在网络中接收到DHCPdiscover发现信息的DHCP服务器都会做出响应，它从尚未出租的IP地址中挑选一个分配给DHCP客户机，向DHCP客户机发送一个包含出租的IP地址和其他设置的DHCPoffer信息。
- 客户端收到了DHCP Server的DHCP Offer 报文之后，则向DHCP Server发送所需要的IP地址请求，该信息中包含向它所选定的DHCP服务器请求IP地址的内容。因为有多个DHCP 服务器，所以客户端是以广播方式发送DHCP Request报文，还因为要通知所有的DHCP服务器，他将选择某台DHCP服务器所提供的IP地址 。
- 服务器收到请求之后，给客户端发送ACK响应报文。
- 以后DHCP客户端每次重新登录网络时，就不需要再发送DHCPdiscover发现信息了，而是直接发送包含前一次所分配的IP地址的DHCPrequest请求信息。

更多资料获取：http://learning.huawei.com/cn

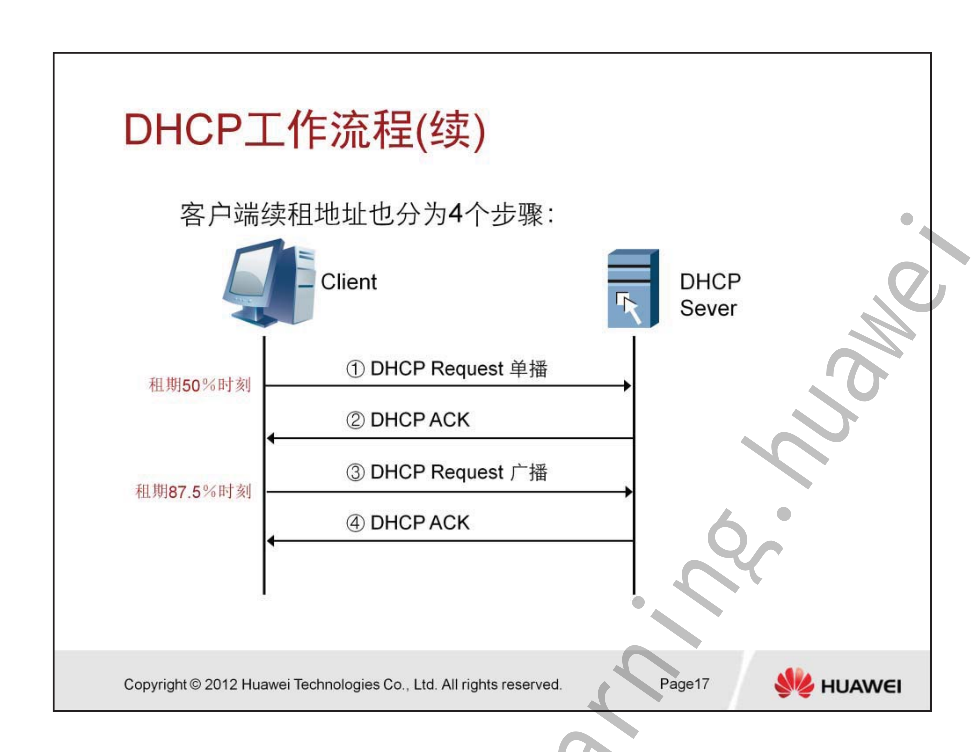
当DHCP服务器收到这一信息后，它会尝试让DHCP客户机继续使用原来的IP地址，并回答一个DHCPack确认信息。如果此IP地址已无法再分配给原来的DHCP客户机使用（比如此IP地址已分配给其它DHCP客户机使用），则DHCP服务器给DHCP客户机回答一个DHCPnack否认信息。当原来的DHCP客户机收到此DHCPnack否认信息后，它就必须重新发送DHCPdiscover发现信息来请求新的IP地址。

相对于DHCP CLIENT，DHCP SERVER的行为比较简单,DHCP SERVER的行为完全由DHCP CLIENT来驱动，只需根据收到的DHCP CLIENT的各种请求报文，相应的响应不同的DHCP 响应报文即可。

DHCP SERVER给DHCP CLIENT分配IP地址原则：DHCP SERVER收到DHCP请求报文后，将会首先查看“giaddr （gateway IP address)”字段是否为0 ，如果不为0，则就会根据此IP地址所在网段从相应地址池中为CLIENT分配IP地址，如果为0，则DHCP SERVER认为CLIENT与自己在同一子网中，将会根据自己的IP地址所在网段从相应地址池中为CLIENT分配IP地址。

DHCP SERVER还能实现地址池管理功能。

更多资料获取：http://learning.huawei.com/cn



客户端获得的IP地址都带有一个租期，在租期满了之后，如果没有续约行为，则用户的IP地址会被服务器收回。

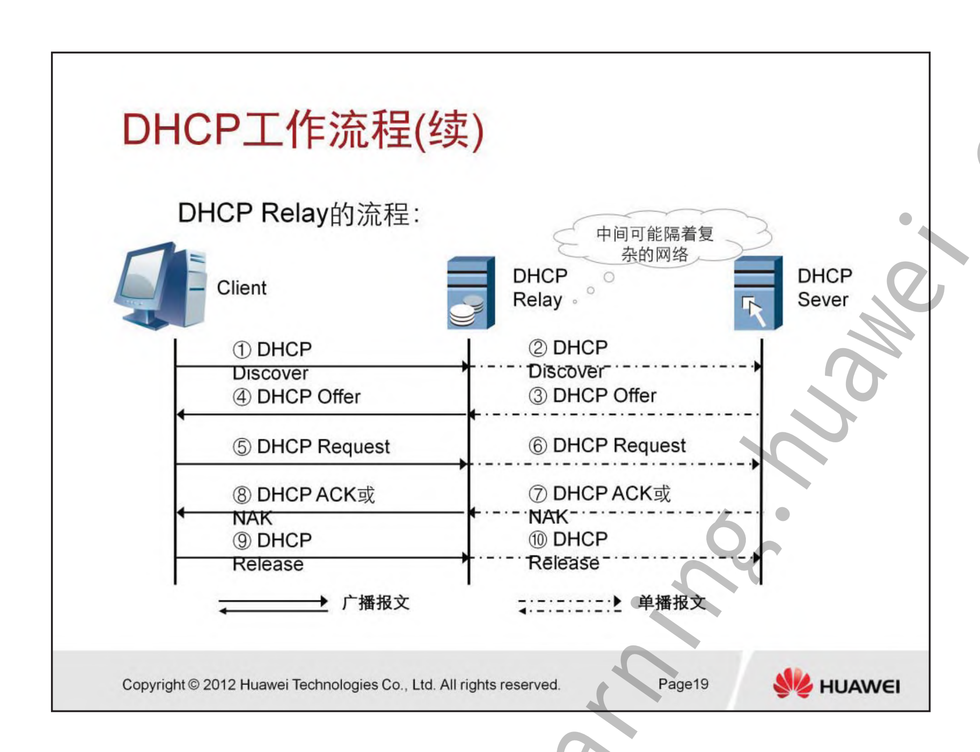
流程如下：

- 当租期到50%的时候，客户端发送DHCP Request 报文，进行续约，因为前期已经有DHCP Server的信息，故此时是单播。
- 服务器收到DHCP Request之后，给客户端发送响应信息，此次重置租期。
- 如果服务器在租期50%的时候没有收到客户的续约报文(DHCP Request)，则客户的IP地址租约不会重置。到租约时间的87.5%的时候，客户端还没有收到服务器的响应（针对客户端50% 发出的请求），客户端会假定原来的DHCP服务器不可用，并开始发送广播的DHCPREQUEST报文。网络上的任何DHCP服务器均可以响应此客户端的请求，并向此客户端发送DHCPACK报文或者DHCPNAK报文。
    - 如果客户端收到一个DHCPACK报文，那么就返回到绑定状态，且重新设置“租期更新和重绑定定时器”。
    - 如果客户端收到的都是DHCPNAK报文，那么就返回到初始化状态。此时客户端必须立即停止使用此IP地址，

更多资料获取：http://learning.huawei.com/cn

  - 并且返回到初始化状态，重新申请新的IP地址。
- 若客户端在“到达租期定时器”到期前都没有收到响应，客户端必须立即停止使用此IP地址，并且返回到初始化状态，重新申请新的IP地址。
- DHCP 服务器收到续约请求后，发送DHCP ACK响应用户的续约，此时重置用户的租期。

更多资料获取：http://learning.huawei.com/cn



DHCP中继的应用环境:

- 早期的DHCP协议只适用于DHCP客户端和服务器处于同一个网段内的情况，不能跨网段，早期的DHCP报文格式中少很多字段，如giaddr、hop等，后期进行了扩充。因此，为进行动态主机配置，每个网段需要一个DHCP服务器，这显然是很不经济的。
- DHCP中继（DHCP Relay）功能的引入解决了这一难题：客户端可以通过DHCP中继与其他网段的DHCP服务器通信，最终取得合法的IP地址。这样，多个网段的DHCP客户端可以使用同一个DHCP服务器，既节省了成本，又便于进行集中管理。

DHCP报文多采用广播方式，是无法穿越多个子网的，当DHCP报文要穿越多个子网时，就要有DHCP RELAY的存在。

DHCP RELAY可以是路由器，也可以是一台主机，总之，DHCP RELAY要监听UDP目的端口号为67的所有报文。

当DHCP RELAY收到一个这样的报文时，会首先判断是否是用户的请求报文，如果是而且“giaddr”（gateway IP address)字段为0，则把自己的IP地址填入此字段，并把此报文单播给真正的DHCP SERVER，以实现DHCP报文穿越多个子网的目的。

当DHCP RELAY发现这是DHCP SERVER的响应报文时，

更多资料获取：http://learning.huawei.com/cn

会根据 “flag”字段中的广播标志位来广播或单播封装好的报文后，传送给DHCP CLIENT。

DHCP Relay场景下的流程如下：

- 客户端发送DHCP Discover广播报文在网络上寻找DHCP Server。
- DHCP Relay收到DHCP Discover报文后，把自己的IP地址填入giaddr 字段，发送给DHCP Server。并把DHCP Server 字段填充相应的Server IP 地址，发送单播报文给DHCP Server。
- DHCP Server 收到 DHCP Relay发来的Discover报文，给DHCP Relay相应，给客户端响应DHCP Offer报文，其中带有DHCP服务器的IP地址。
- DHCP Relay 把收到的Offer报文传送给客户端。
- 客户此时已经找到DHCP Server，开始发送DHCP Request请求IP地址。
- DHCP Relay 收到此请求后，把giaddr字段填充自己的IP地址，并把DHCP Request报文单播发给DHCP Server。
- Server收到请求后，给客户端分配IP地址，发送DHCP ACK 报文，如果客户的请求地址无效则发送NAK报文给DHCP Relay。Server发送给DHCP Relay的报文中，Your IP address填充分配给客户的IP地址。
- DHCP Relay收到服务器的响应报文，把相应的报文发送给客户端。
- 如果用户想释放IP地址，则发送DHCP Release报文，报文中Client IP Address填充已获得的IP地址。
- DHCP Relay收到Release报文后，把giaddr字段填充自己的IP地址发送给Server。Server收到请求后，释放先前分配的IP地址。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

DHCP Snooping

DHCP Snooping原理

- DHCP Snooping 是一种DHCP 安全特性，通过截获DHCP Client 和DHCP Relay之间的DHCP报文并进行分析处理，可以过滤不信任的DHCP 报文并建立和维护一个DHCP Snooping 绑定表。
  - 绑定表包括MAC地址、IP地址、租约时间、VLAN ID、接口信息。
- DHCP Snooping通过对这个绑定表的维护，建立一道在DHCP Client和DHCP Server之间的防火墙。
- DHCP Snooping可以解决设备应用DHCP时遇到的DHCP DoS（Denial of Service）攻击、DHCP Server仿冒攻击、DHCP仿冒续租报文攻击等问题。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page22 

DHCP SNOOPING记录一个DHCP的绑定表，绑定表记录用户的MAC、IP、VLAN、端口等信息，在用户上线的时候，创建该用户的绑定表，当用户下线时删除此绑定表。

在绑定表生成的过程中或生成后，通过对收到的DHCP报文的检查，将报文中的对应字段与DHCP绑定表进行比较，发现DHCP攻击，然后丢弃该报文，以阻止DHCP攻击。

另外，DHCP SNOOPING区分信任端口和非信任端口，把通向DHCP Server（运营商网络内部）的接口设成“信任（Trusted）”，其它接口（连接运营商网络外部的接口）都设为“不信任（Untrusted）。只有DHCP RELAY设备才添加giaddr标识为RALAY的IP，对非信任端口，不处理DHCP Reply报文即带giaddr为非0的DHCP报文，以防止DHCP Server仿冒攻击。

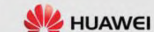
同时，依赖于DHCP SNOOPING表项，可以防止IP欺骗和ARP欺骗等。

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping(续)

DHCP Snooping关键技术

- 信任 / 非信任端口：一般通向DHCP服务器 (运营商网络内部)的端口设成“信任(Trusted)”，其它端口(连接运营商网络外部的端口)都设为“不信任(Untrusted)”。
- 绑定表：建立MAC + IP + VLAN + Port的绑定关系。
- Option82：是DHCP协议报文中选项部分之中的一项，用于记录报文入端口类型，端口号，VLAN信息以及桥MAC地址，是生成绑定表的重要部分。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page23 

发往信任区的DHCP报文中的giaddr为非0，收到非信任区的DCHP报文的giaddr字段为0。

DHCP Snooping 主要是解决网络中应用DHCP时设备遇到DHCP DoS攻击、DHCP Server仿冒攻击、ARP中间人攻击及IP/MAC Spoofing攻击的问题。

DHCP Snooping既可以应用在二层网络设备上，又可应用在三层网络设备上。

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping(续)

DHCP Snooping绑定表分为动态绑定表和静态绑定表。

- 静态绑定表：

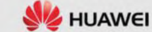
  按照实际需求在报文入端口手工输入，没有租期限制

  - 用途：一些重要设备（如服务器）和一些高端用户需采用静态方式，一是没有租期限制，二是安全性高且便于管理。

- 动态绑定表：

  DHCP客户端在申请IP地址过程中，根据DHCP报文内容在报文入端口自动生成，存在老化时间，有租期限制。

  - 用途：生成方便，常用于非重要设备。不过绑定表存在老化时间，且不便于管理。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page24 

静态绑定：

- 如果分配给用户的IP为静态IP地址，那么可以配置这些分配的IP地址的静态绑定表项，防止用户盗用静态IP地址。但是如果静态IP用户比较多，就需要一一配置。否则无法隔离盗用静态IP地址的非法用户。
- 对于静态分配给用户的IP地址，设备不会自动学习用户的MAC地址，也不能建立绑定关系表，所以需要手动建立绑定关系表。

动态绑定：

- DHCP Snooping绑定表中的动态表项不需配置，使能DHCP Snooping功能即可自动生成。
- 对于动态分配给用户的IP地址，设备会自动学习用户的MAC地址并建立绑定关系表，此时不需要配置绑定表。

更多资料获取：http://learning.huawei.com/cn



RFC3046（DHCP Relay Agent Information Option），提出了Option82的应用，由DHCPRelayAgent插入到用户的DHCP报文，DHCP服务器通过识别Option82来执行IP地址分配策略或其它策略。DHCP服务器的响应报文也带有Option82，Relay Agent将Option82剥离后发给用户。Agent Information Field中包括多个子选项，每个子选项格式为SubOpt/Length/Value三元组。目前定义了两个子选项:

1 Agent Circuit ID Sub-option：用于标识用户电路。

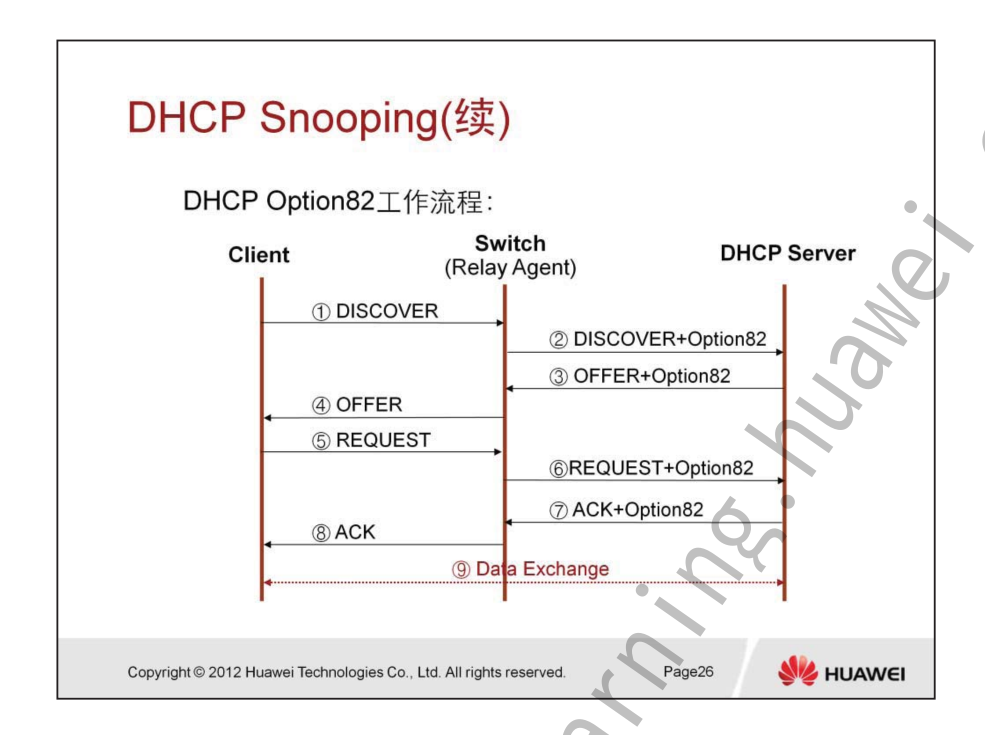
2 Agent Remote ID Sub-option：用于标识电路端点的远端主机。

当在DHCP Relay上应用Option82功能时，如果用户构造的Option82中没有包含接口相关信息，那么生成的绑定表中也不包含接口信息。这可能导致:

- Server的DHCP Reply报文会被同一VLAN内其他接口下的用户监听到。
- 用户上线后，如果同一VLAN内其他接口下的用户伪造IP地址和MAC地址，就可以仿冒此合法用户。

在二层上应用DHCP Snooping时，不配置Option82功能也可以获得绑定表所需的接口信息。

更多资料获取：http://learning.huawei.com/cn



客户端发送DHCP Discover报文在网络上寻找DHCP服务器。

Switch收到用户的DHCP Discover报文后，DHCP Relay Agent给用户的DHCP报文加上Option82信息，后发送给DHCP Server。

DHCP Server收到Discover报文后给客户相应，此时响应的DHCP Offer报文带有先前的Option82信息。

Switch收到DHCP Server的响应信息后，DHCP Relay Agent去掉Option82信息，并把此DHCP Offer报文传送给客户端。

此时客户已经发现了DHCP Server，开始发送DHCP Request报文进行IP地址请求。

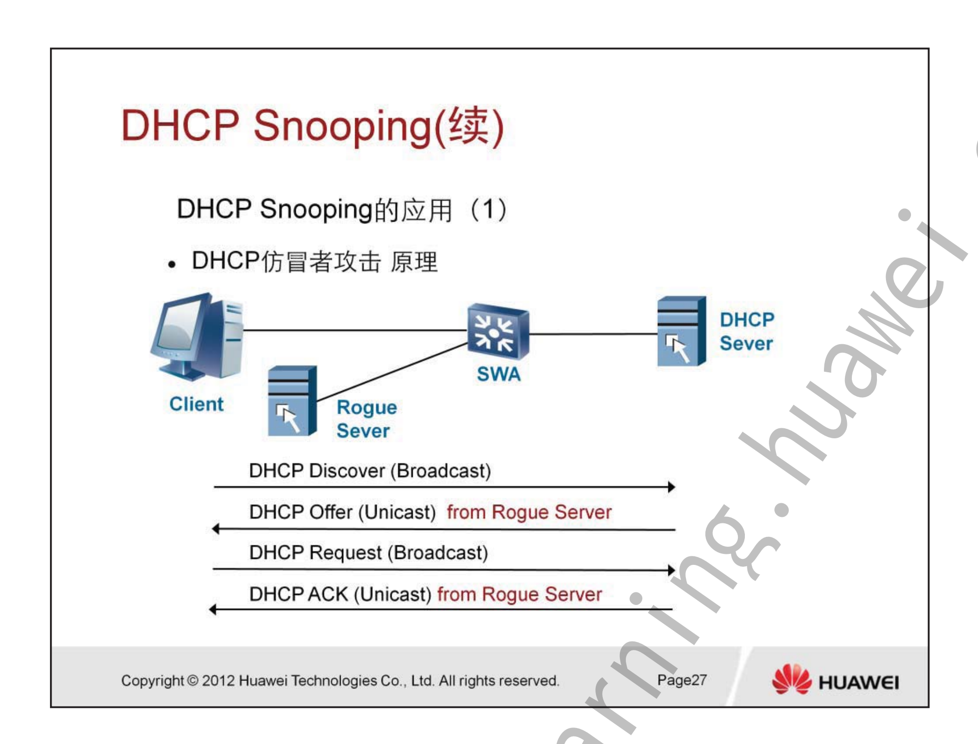
Switch收到客户的DHCP Request报文后，DHCP Relay Agent给报文打上Option82信息，发送给DHCP Server。

DHCP Server收到带有Option82信息的DHCP Request报文后，开始给客户分配IP地址，如果客户申请的IP地址无效或者过期则发送DHCP NAK，如果满足条件则发送DHCP ACK 响应客户。此时的ACK报文带有Option82信息。

Switch收到DHCP 服务器带有Option82信息的ACK响应报文后，DHCP Relay Agent去掉Option82信息，把DHCP ACK信息发送给客户端。

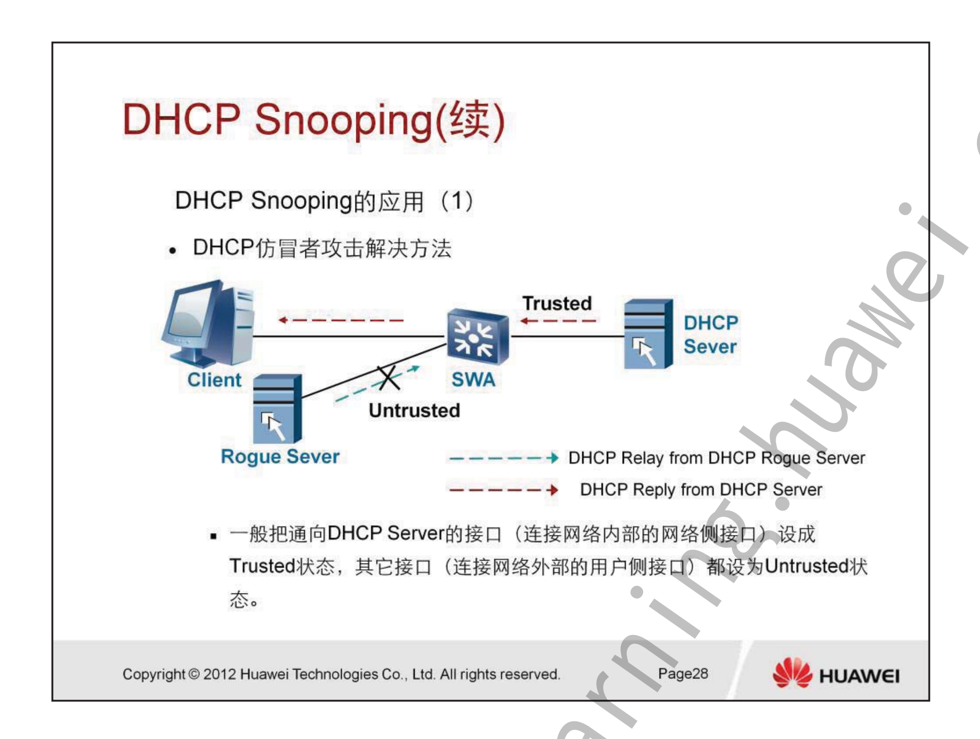
当地址申请结束后，用户就可以进行数据交换了。

更多资料获取：http://learning.huawei.com/cn



由于DHCP请求报文以广播形式发送，所以DHCP Server仿冒者可以侦听到，并且回应仿冒信息，可以通过回应错误的网关、DNS服务器、IP等，达到DoS(拒绝服务)的目的。

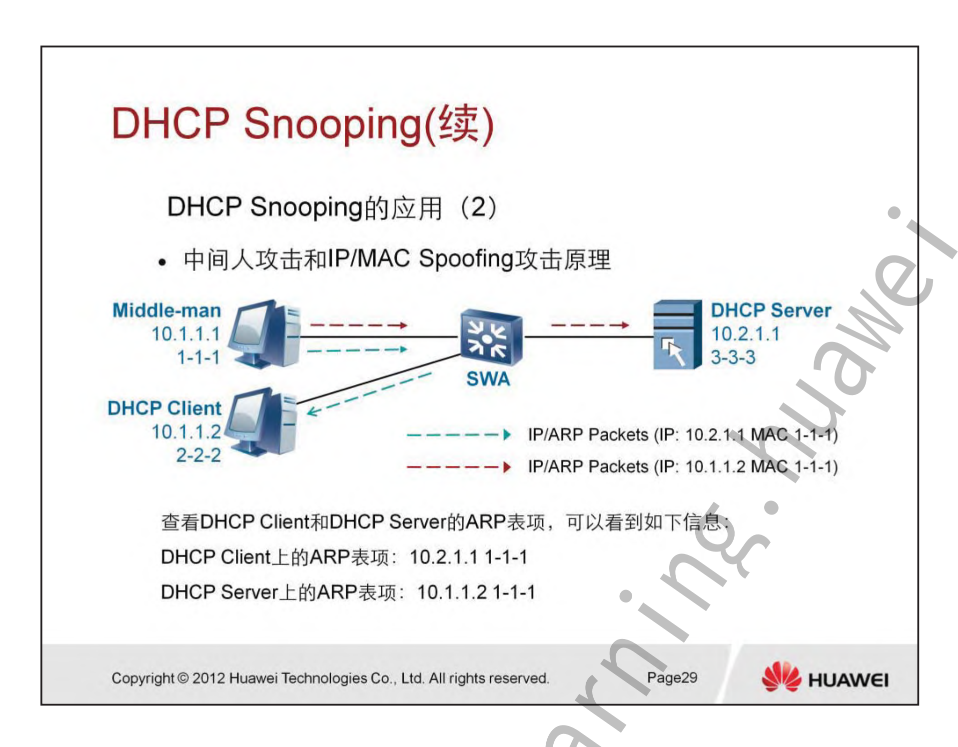
更多资料获取：http://learning.huawei.com/cn



为防止DHCP Server仿冒者攻击，可配置Trusted/Untrusted接口。

把某个物理接口或者VLAN 设置为Trusted或者Untrusted状态。凡是从Untrusted接口上收到的DHCP ReplyOffer、ACK、NAK）报文直接丢弃，这样可以隔离DHCP Server仿冒者攻击。

更多资料获取：http://learning.huawei.com/cn



如图所示，Middle-Man 通过发送IP 或ARP 报文，让DHCP Server学到DHCP Client的IP地址10.1.1.2 和自己的MAC 地址1-1-1。在DHCP Server看来，所有的报文都是来自或者发往DHCP Client，而实际上所有的报文都经过Middle-Man 处理。

Middle-Man 再发送IP 或ARP 报文，让DHCP Client 学到DHCP Server的IP地址10.2.1.1 和自己的MAC 地址1-1-1。在DHCP Client 看来，所有的报文也都是来自或者发往DHCP Server，而实际上所有的报文都经过Middle-Man 处理。

这样Middle-Man就可以达到仿冒DHCP Server和DHCP Client 的目的，从而获得DHCP Server和DHCP Client之间交互的信息。

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping(续)

DHCP Snooping的应用（2）

- 中间人攻击和IP/MAC Spoofing攻击解决方法

为了防止中间人攻击或IP/MAC Spoofing 攻击，可以在交换机上配置DHCP Snooping 功能，使能DHCP Snooping 绑定表功能后，只有接收到的报文的信息和绑定表中的内容一致才会被转发，否则报文将被丢弃。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page30 HUAWEI

接口接收到ARP或者IP报文，使用ARP 或者IP 报文中的“源IP+源MAC”匹配DHCP Snooping 绑定表，如果配置强策略，则匹配不到绑定表就丢弃。

对于配置静态IP 的用户，由于没有通过DHCP请求而获得IP，所以，没有对应的DHCP Snooping绑定表项，该用户发出的ARP、IP 报文会被丢弃，从而防止该用户非法使用网络。只能通过配置静态DHCP Snooping 绑定表来允许静态IP用户访问网络。

对于盗用其他合法用户IP地址的用户，同样由于不是自己通过DHCP请求而获得IP的，IP对应的DHCP Snooping绑定表项中的MAC以及接口与盗用者的不一致，盗用者发出的ARP、IP报文会被丢弃，从而防止该盗用者非法使用网络。

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping(续)

DHCP Snooping的应用（3）

- 饿死攻击
  - 攻击原理：在饿死攻击方式中，攻击者不断变换物理地址，尝试申请地址池中所有的IP 地址，直到耗尽DHCP Server 地址池中的地址，导致其他正常用户无法获得地址。
  - 解决方案：通过MAC 地址限制功能可以防止饿死攻击。通过限制交换机接口上允许学习到的最多MAC 地址数目，防止用户通过变换MAC 地址，大量发送DHCP 请求，同时也限制了一个接口上的用户数目。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page31



MAC地址限制功能一般部署在二层设备上。通过限制交换机接口上允许学习到的最多MAC地址数目，防止用户通过变换MAC地址，大量发送DHCP请求，同时也限制了一个接口上的用户数目。

QinQ方式下的MAC地址限制：

- 在实际的应用中，用户通过DSLAM（Digital Subscriber Line Access Multiplexer）设备接入网络，通过给不同的用户配置不同的VLAN 来隔离用户，同时为了规避VLAN 总数（4094）的限制，使用QinQ 特性，给用户报文打两层Tag标签，这时如果在网关上部署了MAC地址限制功能，就需要能够基于两层Tag对MAC地址数目进行限制。

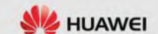
更多资料获取：http://learning.huawei.com/cn



DHCP Snooping(续)

DHCP Snooping的应用（4）

- 改变CHADDR 值的饿死攻击
  - 攻击原理：在这种攻击方式中，如果攻击者改变的不是数据帧头部的源MAC，而是改变DHCP 报文中的CHADDR（Client Hardware Address）值来不断申请IP 地址，而交换机仅根据数据帧头部的源MAC 来判断该报文是否合法，那么MAC 地址限制方案不能起作用。
  - 解决方案：可以使用DHCP Snooping 检查DHCP REQUEST 报文中CHADDR 字段的功能。如果该字段跟数据帧头部的源MAC 相匹配，便转发报文；否则，丢弃报文。

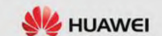
Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page32 

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping(续)

DHCP Snooping的应用（5）

- ARP攻击原理
  - ARP攻击方式：有针对主机的，也有针对网关的；有地址欺骗型的，也有野蛮攻击型的；有来自病毒的攻击，也有来自使用非法软件的人为攻击。
  - ARP攻击根源：ARP协议本身过于简单和开放，没有任何的安全手段。
  - ARP攻击危害：ARP地址欺骗攻击一般针对个别或一定范围内的主机进行，危害相对较小。但针对网关设备的大流量ARP DDOS攻击，由于其网络位置的特殊性，将造成大面积用户“掉线”。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page33 

更多资料获取：http://learning.huawei.com/cn



DHCP Snooping(续)

DHCP Snooping的应用（5）

- ARP攻击 解决方法
  - 在应用DHCP服务器的组网环境下， 建立可信端口（trust Port），通过监控可信端口的DHCP 报文获得IP/MAC地址绑定表，这是DHCP Check IP/ARP安全检查手段的重要依据。实际上也是一种安全焦点的转移，把ARP安全问题转换为别的安全问题。
  - DHCP Snooping Chek IP/ARP 依据可信端口上生成的绑定表，过滤掉所有不匹配的IP/ARP报文。大大的提高了防攻击的能力。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page34 

在应用DHCP服务器的组网环境下，DHCP Snooping 防攻击的效果比较好。其原因主要是把安全目标转移了，从没有任何安全手段的ARP协议转到DHCP 协议。由于DHC 服务器的应用环境比主机应用环境要好得多，所以，可以建立可信端口(trust Port)，通过监控可信端口的DHCP报文获得IP/MAC地址绑定表，这是DHCP Check IP/ARP安全检查手段的重要依据。这实际上也是一种安全焦点的转移，把ARP安全问题转换为别的安全问题。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



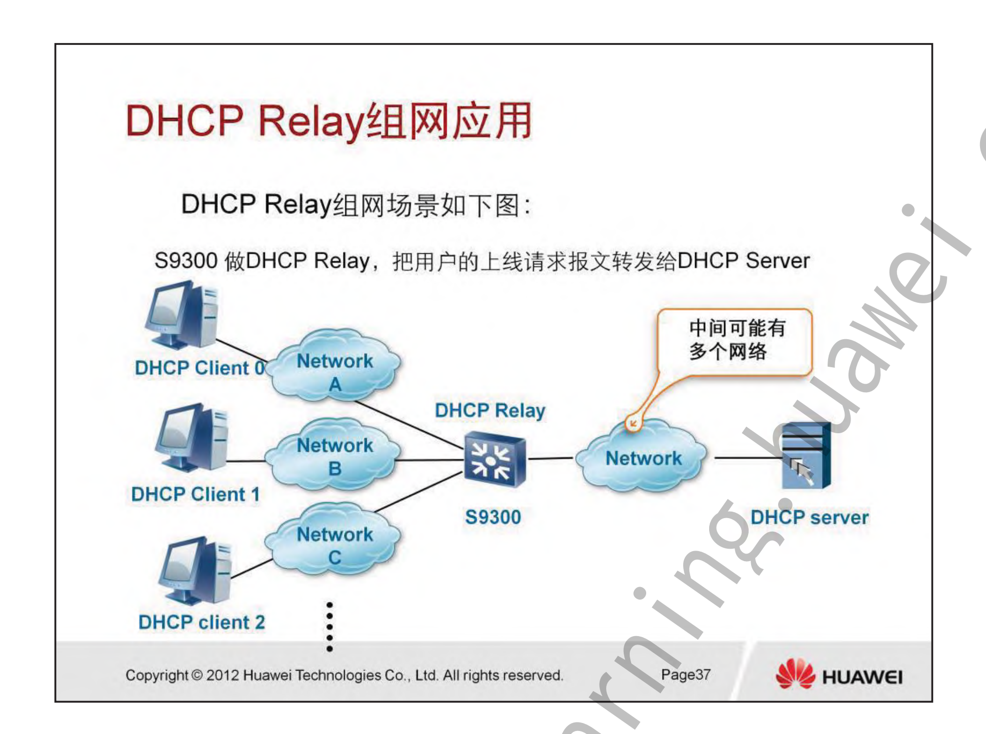
DHCP 在S9300上的配置

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page36



更多资料获取：http://learning.huawei.com/cn



使用DHCP Relay的应用环境：

- 本地网络没有配置DHCP 服务器，可以在S9300 上启动DHCP中继功能，从客户端来的DHCP请求可通过DHCP中继传到DHCP服务器。此时，为了保证客户端正常获取IP地址，该服务器必须是基于全局地址池的DHCP 服务器，即服务器与DHCP中继相连的接口不允许再配置接口地址池。

接口地址池优先于全局地址池分配地址。即若接口上存在接口地址池，即使全局地址池已存在，客户端优先从接口地址池中获取地址。

更多资料获取：http://learning.huawei.com/cn

DHCP Relay组网应用(续)

S9300上的DHCP Relay配置：

- dhcp server group dhcpgrp

  #配置DHCP服务器组的组名
- dhcp-server 10.1.1.1 24

  #配置DHCP服务器组中的DHCP服务器IP地址
- interface vlanif 10
- ip address 192.168.1.1 24

  #配置启动DHCP Relay功能的接口编号及接口的IP地址
- dhcp select relay
- dhcp relay server-select dhcpgrp

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page38 

DHCP服务器组全局最多可以配置64个。

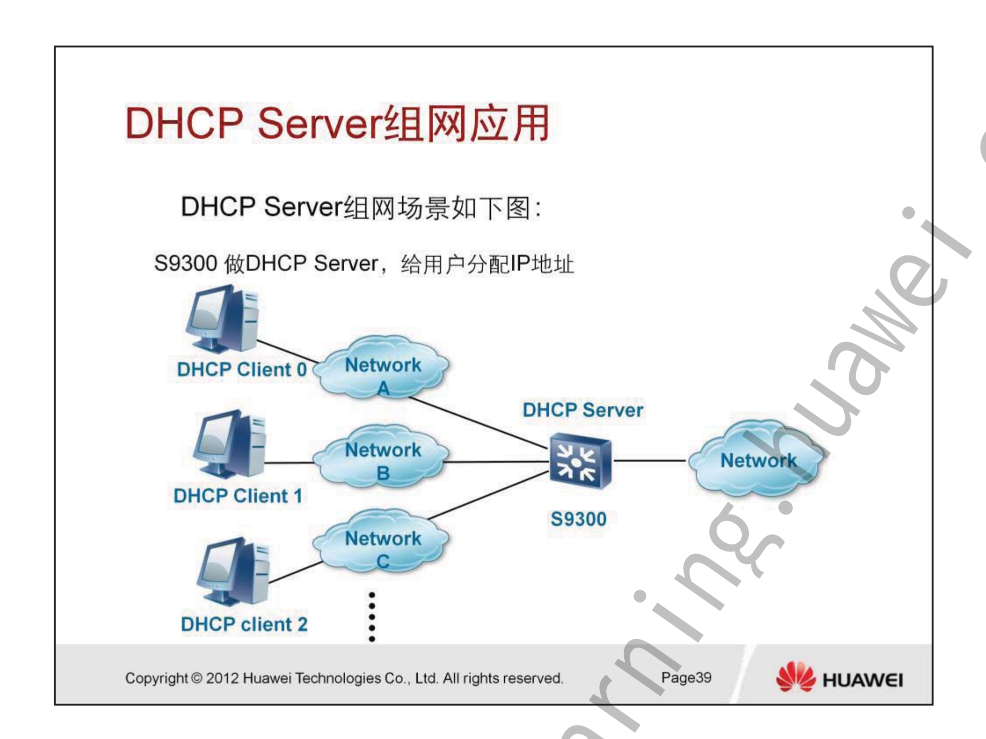
每个DHCP服务器组下最多可以配置8个DHCP服务器。不指定索引时，系统将自动分配一个空闲的索引。

在S9300 的VLANIF接口下使能DHCP中继功能，该接口对收到的DHCP报文将进行中继转发。

- DHCP服务器和DHCP客户端之间的DHCP报文中继次数不能超过4次，否则DHCP报文将被丢弃。
- 一个super-Vlan下使能了DHCP中继功能后，则该super-Vlan下不能使能DHCP Snooping功能。

一个DHCP服务器组可以对应多个VLANIF 接口。一个VLANIF接口下只能指定一个DHCP服务器组，即一个VLANIF接口下的DHCP request 报文最多可以中继转发到一个DHCP 服务器。

更多资料获取：http://learning.huawei.com/cn



在分布式网络中，业务的部署尽量靠近用户。S9300 做为汇聚节点，通过本地DHCP Server或远端DHCP Server 为用户分配IP 地址，使得客户端在S9300 上终结。

更多资料获取：http://learning.huawei.com/cn

DHCP Server组网应用(续)

S9300上的DHCP Server配置：

- `dhcp enable`

  #使能DHCP功能

- `interface vlanif 10`

  ```
  ip address 192.168.1.1 24
  dhcp select global
  ```

  #配置基于全局地址池的DHCP服务器。

  （采用全局地址池的DHCP服务器模式时，还需要配置全局地址池。）

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page40 

S9300 作为DHCP 服务器时，可以设置接口对目的地址是本机的DHCP 报文的处理模式。S9300 可以通过全局地址池来分配地址，也可以通过基于接口的接口地址池来分配地址。

客户端和S9300 要在同一个子网内，在同一接口下只能选择一种配置方式。

更多资料获取：http://learning.huawei.com/cn

DHCP Server组网应用(续)

S9300上的DHCP Server配置(续):

- interface vlanif 10

  ip address 192.168.1.1 24

  dhcp select interface

  #配置基于接口地址池的DHCP服务器

  （接口地址池优先于全局地址池分配地址。）

- dhcp server ping packets 5

  #配置防止IP地址重复分配功能（可选）

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page41 

配置基于全局地址池的 DHCP 服务器，从所有接口上线的用户都可以选择该地址池中的地址。

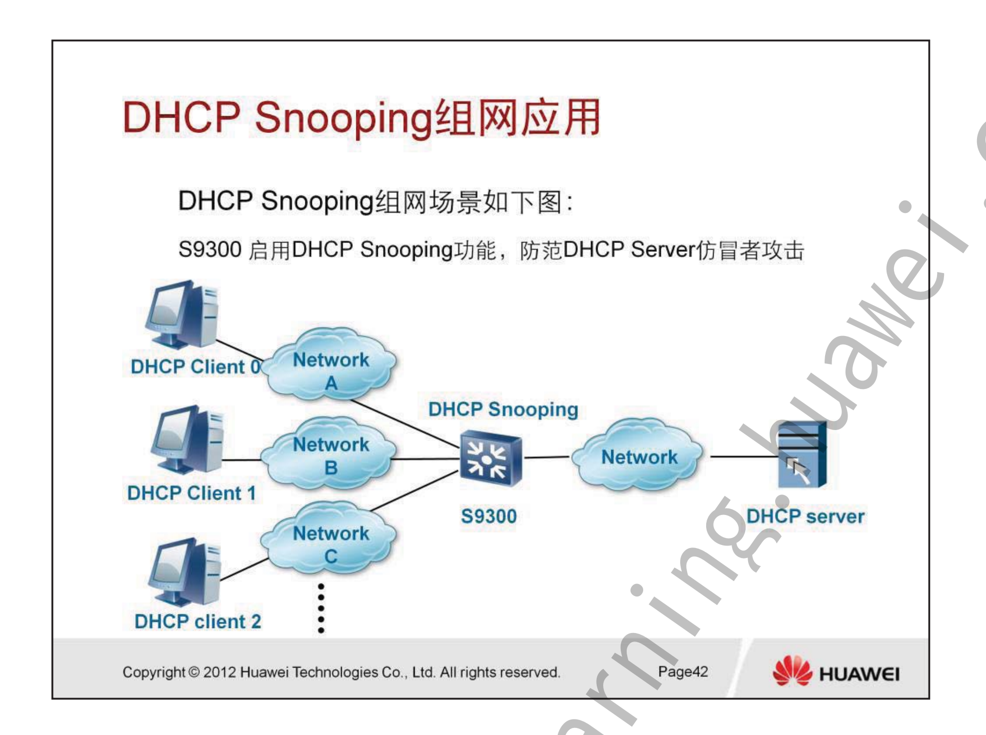
- 全局地址池配置:
  - ip pool 2
  - ip pool 1
  - gateway-list 10.1.1.126
  - network 10.1.1.0 mask 255.255.255.128
  - dns-list 10.1.1.2

配置基于接口地址池的DHCP 服务器，从这个接口上线的用户都从该地址池中分配地址。从此接口上来的用户分配和接口同端的IP地址，网关为该接口的IP。

接口地址池优先于全局地址池分配地址。即若接口上存在接口地址池，即使全局地址池已存在，客户端优先从接口地址池中获取地址。

DHCP服务器通过发送Ping报文探测地址的使用情况，以防止重复分配IP地址导致地址冲突。缺省情况下，DHCP服务器发送ping报文的最大数目为5，每个ping报文的最长等待回应时间为0毫秒。

更多资料获取 http://learning.huawei.com/cn



支持二层的交换功能，也支持三层的路由特性，在两种应用下都支持 DHCP Snooping 功能

S9300 应用在二层网络中或作为DHCP Relay应用时使能DHCP Snooping，可以防止DHCP攻击。在配置上的区别仅是：作为 DHCP Relay时支持ARP 与DHCP 的联动功能。在二层网络中应用时不支持该功能。

更多资料获取：http://learning.huawei.com/cn

DHCP Snooping组网应用(续)

S9300上的DHCP Snooping配置（防DHCP Server仿冒者攻击）：

- `dhcp enable`

  #使能全局DHCP功能
- `dhcp snooping enable`

  #使能接口DHCP Option82及DHCP Snooping功能
- ```
  interface GigabitEthernet3/0/0
  dhcp option82 insert enable
  dhcp snooping enable
  ```
  #使能接口或VLAN的DHCP Snooping功能
- ```
  interface GigabitEthernet3/0/0
   dhcp snooping trusted
  ```
  #配置接口信任状态

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page43 

使能DHCP Snooping 功能的顺序如下：

- 全局使能DHCP 功能。
- 全局使能DHCP Snooping 功能。
- 在接口或VLAN 下使能DHCP Snooping 功能。

更多资料获取：http://learning.huawei.com/cn



答案:

- DHCP基本原理是主机通过动态的报文交互来获得相应的网络配置和IP地址。
- DHCP基本流程是，首先发送DHCP Discover报文寻找DHCP Server， DHCP Server发送DHCP Offer报文响应，其次主机发送DHCP Request 请求IP地址，Server收到请求后回应ACK响应请求。到50%租期时主机发送DHCP Request 进行续约，如果续约不成功，在87.5%租期时再发送DHCP Request进行续约。
- 见教材P21胶片。
- Option 82 是给用户打上位置信息，以便对实施策略和QoS等。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn

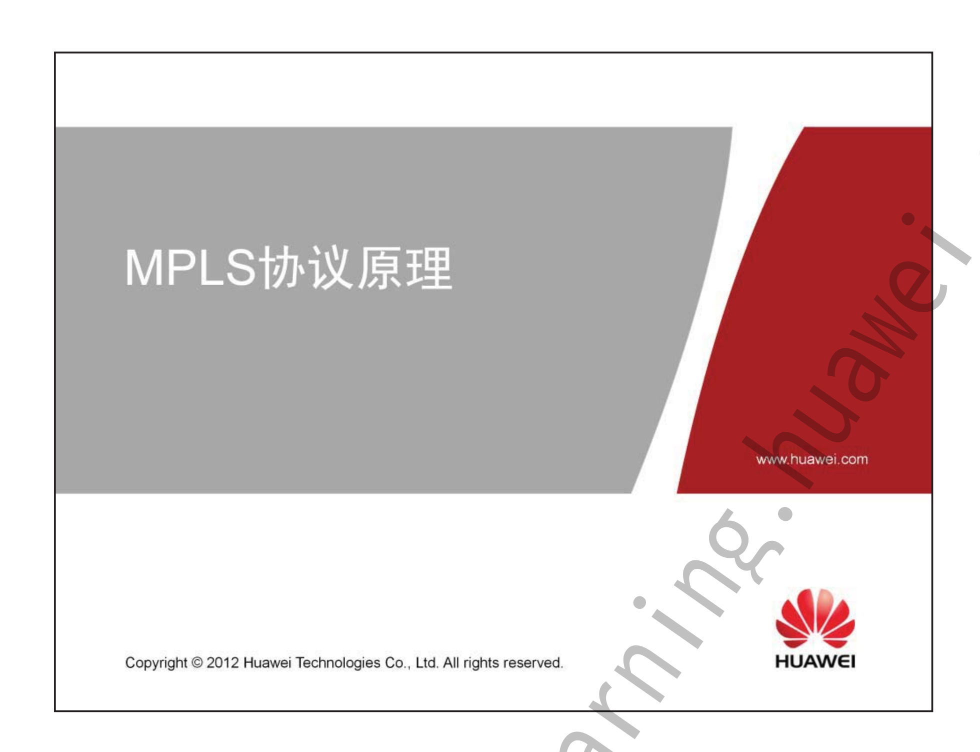
更多资料获取：http://learning.huawei.com/cn

Module 4

MPLS

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

本课程分析了传统IP在转发速率、Qos和流量工程方面的缺陷，介绍了MPLS转发的基本特点。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 描述IP转发流程
- 描述IP转发的缺点
- 解释MPLS转发基本原理
- 描述MPLS应用

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.
Page2
HUAWEI

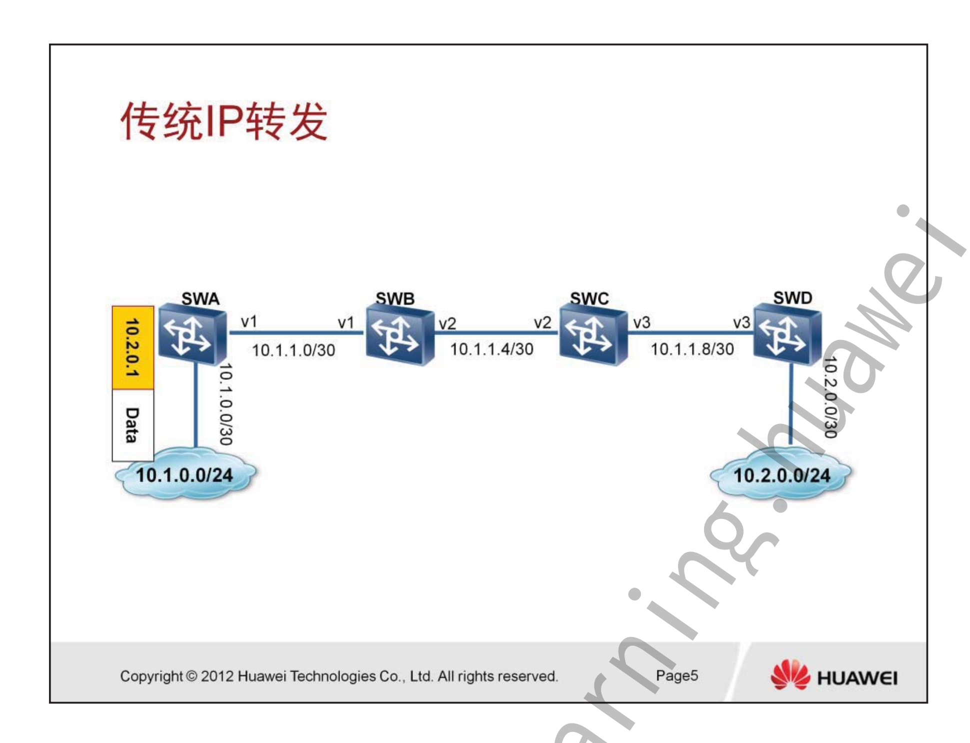
更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



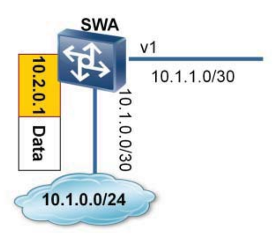
更多资料获取：http://learning.huawei.com/cn



传统的IP转发中，物理层从交换机的一个端口收到一个报文，上送到数据链路层。数据链路层去掉链路层封装，根据报文的协议域上送给相应的网络层。网络层首先看报文是否是送给本机的，若是，去掉网络层封装，上送给它的上层协议。若不是，则根据报文的目的地址查找路由表，若找到路由，将报文送给相应端口的数据链路层，数据链路层封装后，发送报文。若找不到路由，将报文丢弃。传统的IP转发采用的是逐跳转发，数据报文经过每一台交换机，都要执行上述过程（如图中SWA收到目的地址为10.2.0.1的数据包，SWA会依次查找路由表，根据匹配的路由表项的进行转发，SWB、SWC、SWD都会进行类似的处理），所以速度缓慢。并且所有的交换机需要知道全网的路由或者默认路由。另外，由于传统IP转发是面向无连接的，所以无法提供好的Qos保证。

更多资料获取：http://learning.huawei.com/cn

传统IP转发(SWA)



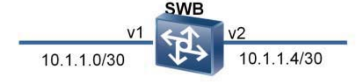
Network	Nexthop
10.1.0.0/24	10.1.0.2
10.1.0.1/32	10.1.0.1
10.1.1.0/30	10.1.1.1
10.1.1.2/32	10.1.1.2
10.1.1.4/30	10.1.1.2
10.1.1.8/30	10.1.1.2
10.2.0.0/24	10.1.1.2

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page6



更多资料获取：http://learning.huawei.com/cn

传统IP转发(SWB)



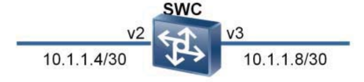
Network	Nexthop
10.1.0.0/24	10.1.1.1
10.1.1.0/30	10.1.1.2
10.1.1.1/32	10.1.1.1
10.1.1.4/30	10.1.1.5
10.1.1.6/32	10.1.1.6
10.1.1.8/30	10.1.1.6
10.2.0.0/24	10.1.1.6

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page7



更多资料获取：http://learning.huawei.com/cn

传统IP转发(SWC)



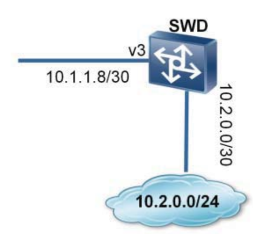
Network	Nexthop
10.1.0.0/24	10.1.1.5
10.1.1.0/30	10.1.1.5
10.1.1.4/30	10.1.1.6
10.1.1.5/32	10.1.1.5
10.1.1.8/30	10.1.1.9
10.1.1.10/32	10.1.1.10
10.2.0.0/24	10.1.1.10

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page8



更多资料获取：http://learning.huawei.com/cn

传统IP转发(SWD)

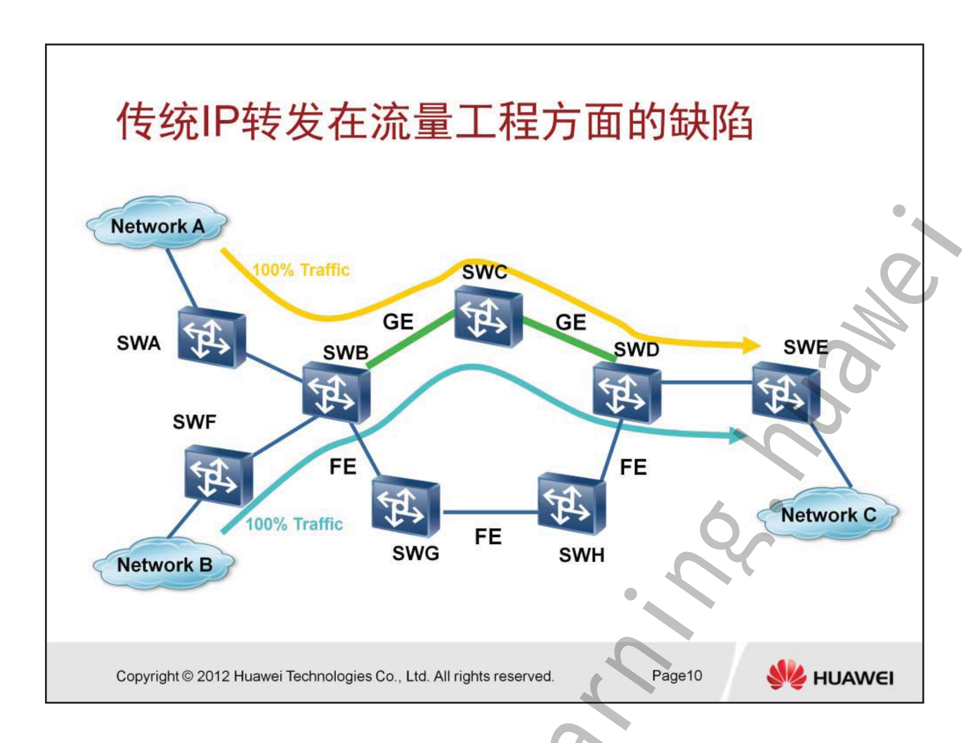


Network	Nexthop
10.1.0.0/24	10.1.1.9
10.1.1.0/30	10.1.1.9
10.1.1.4/30	10.1.1.9
10.1.1.8/30	10.1.1.10
10.1.1.9/32	10.1.1.9
10.2.0.0/24	10.2.0.2
10.2.0.1/32	10.2.0.1



Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page9 HUAWEI

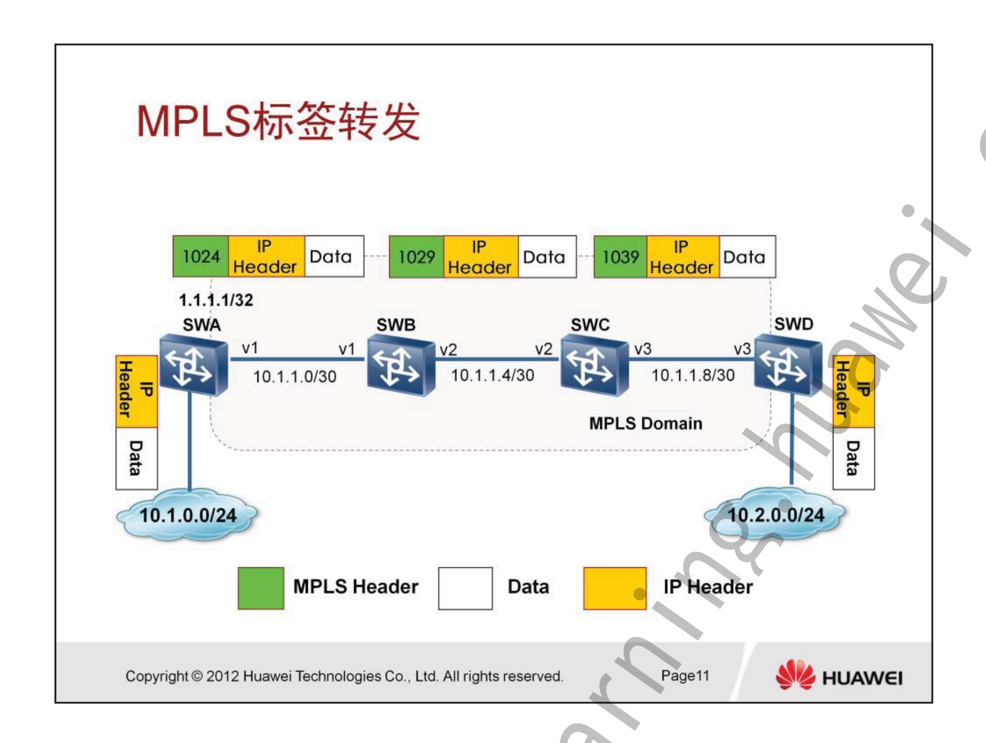
更多资料获取：http://learning.huawei.com/cn



传统IP网络基于IGP Metric计算最优路径，这是远远不够的，往往在现实网络中还需考虑带宽、链路属性等其他因素；基于IP的流量工程是基于IGP面向目的地址的转发，是hop-by-hop的转发，无法实现根据来源来控制流量转发；另外基于IP的流量工程是面向无连接的，不能实现显式路径（Explicit Routing）。

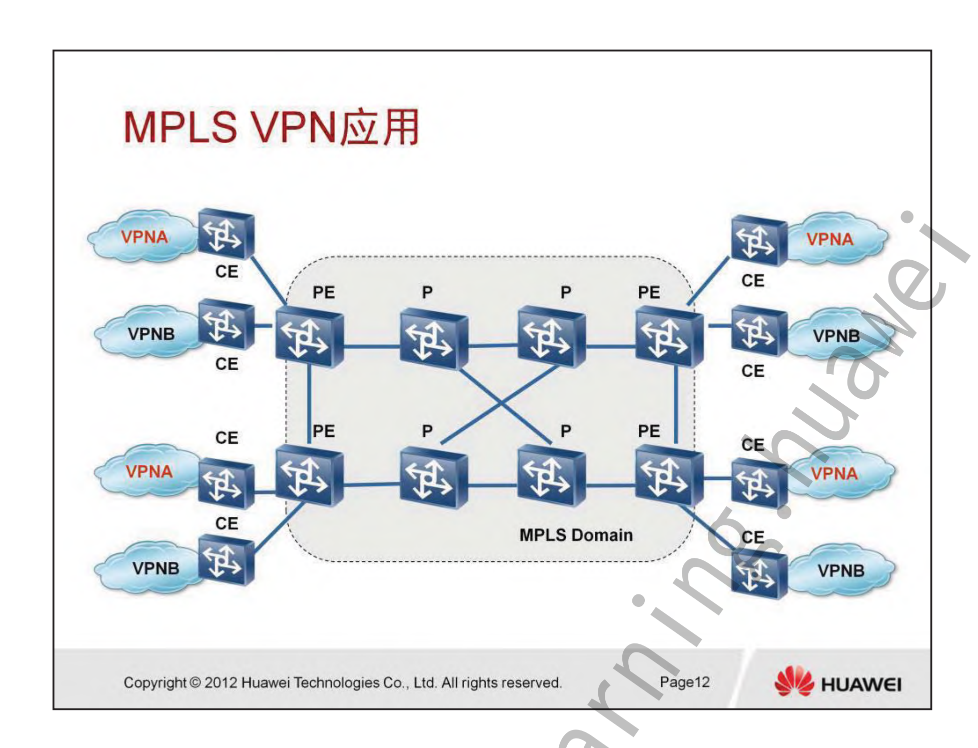
上图中，SWB和SWD之间存在两条路径。传统的IP转发中IGP根据Metric选择最优的路由SWB-SWC-SWD转发所有从Network A和Network B到Network C的IP报文，而SWB-SWG-SWH-SWD链路则闲置，当网络中流量过大，有可能导致最优路径拥塞，但次优路径却空载没有被充分利用。

更多资料获取：http://learning.huawei.com/cn



MPLS是一种标签转发技术，它采用无连接的控制平面和面向连接的数据平面，无连接的控制平面实现路由信息的传递和标签的分发，面向连接的数据平面实现报文在建立的标签转发路径上传送。MPLS域内，交换机不需要查看每个报文的目的IP地址，只需要根据封装在IP头外面的标签进行转发即可（如图中的SWB从SWA收到带有标签的报文，根据标签进行转发，SWC类似）。所以，相对于传统的IP转发，MPLS标签转发大大提高了转发效率。

更多资料获取：http://learning.huawei.com/cn

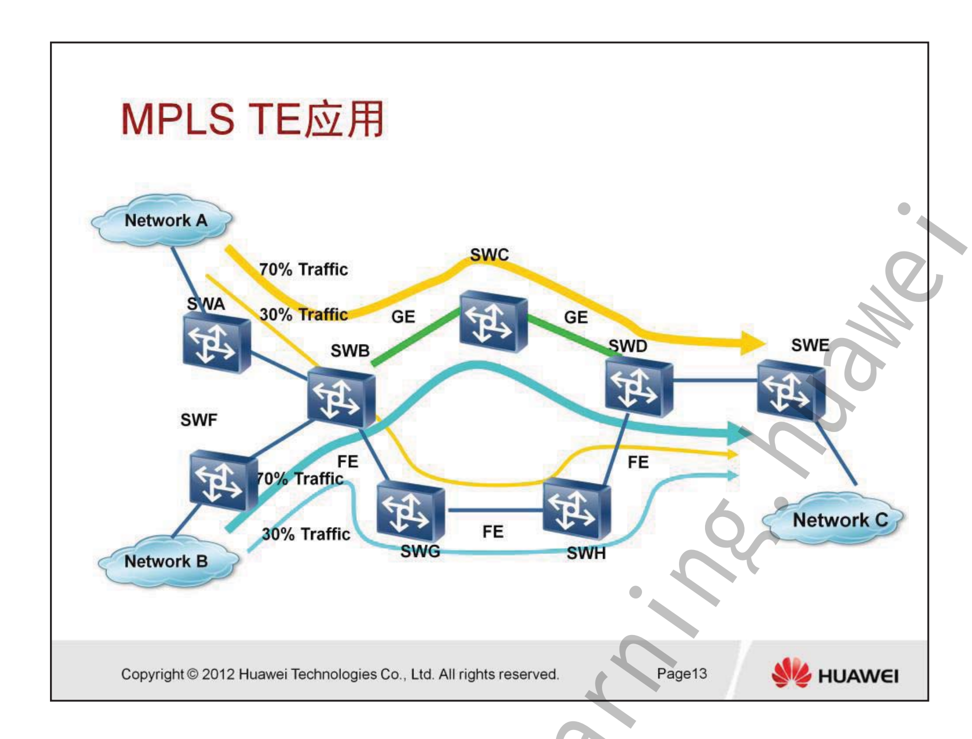


然而，随着ASIC技术的发展，路由查找速度已经不是阻碍网络发展的瓶颈。这使得MPLS在提高转发速度方面不再具备明显的优势。

由于MPLS结合了IP网络强大的三层路由功能和传统二层网络高效的转发机制，在转发平面采用面向连接方式，与现有二层网络转发方式非常相似，这使得MPLS能够很容易地实现IP与ATM、帧中继等二层网络的无缝融合，并为流量工程TE（Traffic Engineering）、虚拟专用网VPN（Virtual Private Network）、服务质量QoS（Quality of Service）等应用提供更好的解决方案。

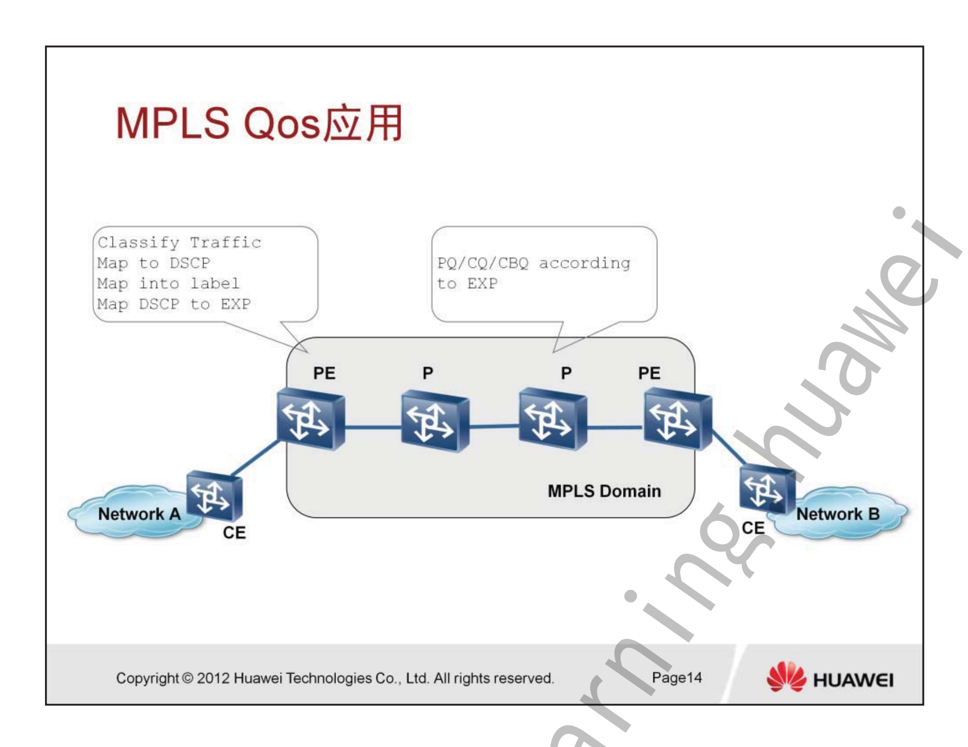
基于MPLS的VPN可以将私有网络的不同分支连接起来，形成一个统一的网络，基于MPLS的VPN还支持对不同VPN间的互通控制。如图所示，CE（Customer Edge）是用户边缘设备；PE（Provider Edge）是服务商边缘交换机，位于骨干网络；P（Provider），是服务提供商网络中的骨干交换机，不与CE直接相连。VPN数据在封装MPLS的标签转发路径中传递。

更多资料获取：http://learning.huawei.com/cn



MPLS TE结合了MPLS技术与TE流量工程，通过建立到达指定路径的LSP隧道进行资源预留，使网络流量绕开拥塞节点，可以达到平衡网络流量的目的。如图所示，Network A到Network C的70%的流量在通过路径SWB-SWC-SWD传递，30%的流量通过SWB-SWG-SWH-SWD传递。Network B到Network C的流量类似。

更多资料获取：http://learning.huawei.com/cn



MPLS和DifferServ完美配合，提供Qos功能。

根据需要在CE上或PE上对业务流进行分类，如可以将DSCP值为2的流分为一类，DSCP值为3的流分为一类，对分类后的流量可以进行流量监管、重新标记EXP等。

PE在给报文加Label（标签）时，把IP报文携带的IP优先级标记映射到标签的EXP域，这样原来由IP携带的服务类型信息，现在由标签携带。

在P交换机和PE交换机之间，根据标签的EXP域，进行有差别的队列调度（如PQ、CQ、CBQ等），即把携带标签的业务流在一条标签转发路径上进行有差别的QoS的传送。

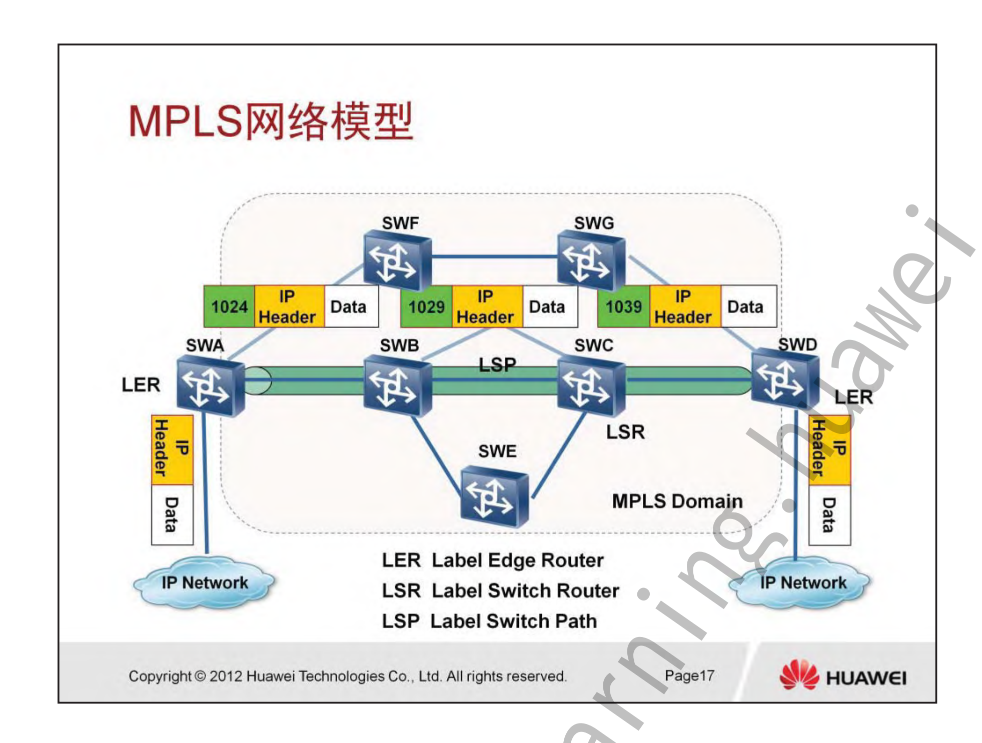
更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



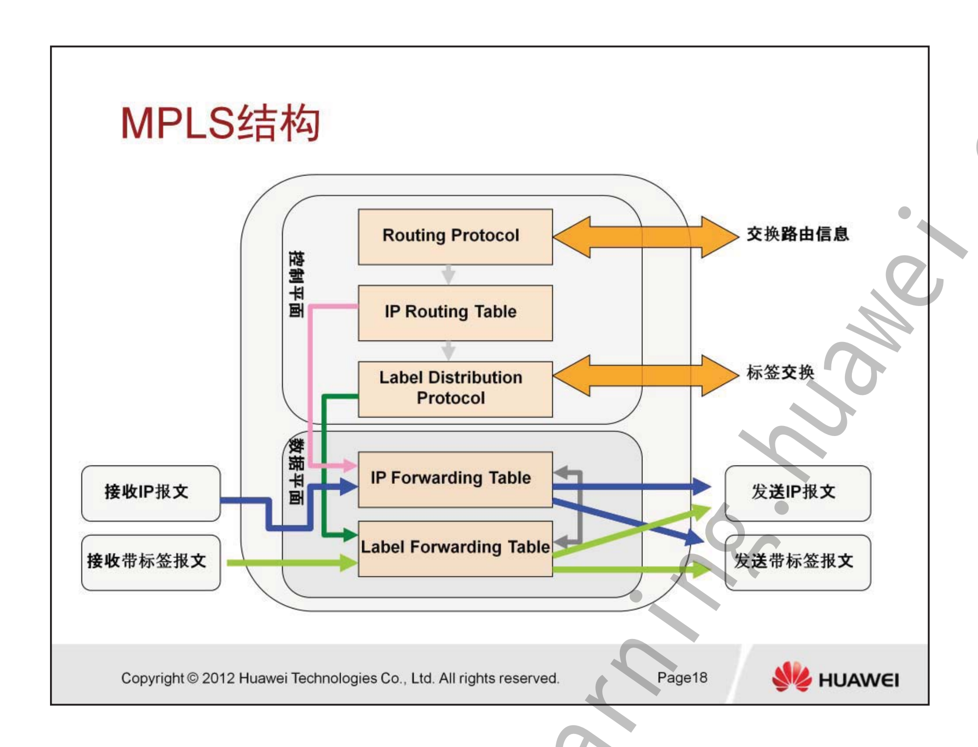
MPLS网络的典型结构如胶片所示：位于MPLS域内的交换机和ATM交换机称为标签交换交换机LSR，位于MPLS域边缘用于连接IP网络或其他非MPLS网络的交换机或ATM交换机称为LER。

在IP网络内进行传统的IP转发，在MPLS域内进行标签转发。

LER和LSR都具有标签转发能力，只是由于两者所处位置不同，对于报文的处理不同。LER负责从IP网络接收IP报文并给报文打上标签，然后送到LSR，反之，也负责从LSR接收带标签的报文并去掉标签然后转发到IP网络；LSR只负责按照标签进行转发即可。

报文在MPLS域内进行转发时经过的路径称为标签转发路径LSP，这条路径是在转发报文之前就已经通过各种协议确定并建立的，报文会在特定的LSP上传递。

更多资料获取：http://learning.huawei.com/cn



MPLS网络根据标签转发报文。那么MPLS中的标签是如何产生的呢？MPLS又是采用什么样的机制实现报文转发的呢？

MPLS包括两个平面：控制平面和数据平面。

控制平面负责产生和维护路由信息以及标签信息。数据平面负责普通IP报文的转发以及带MPLS标签报文的转发。

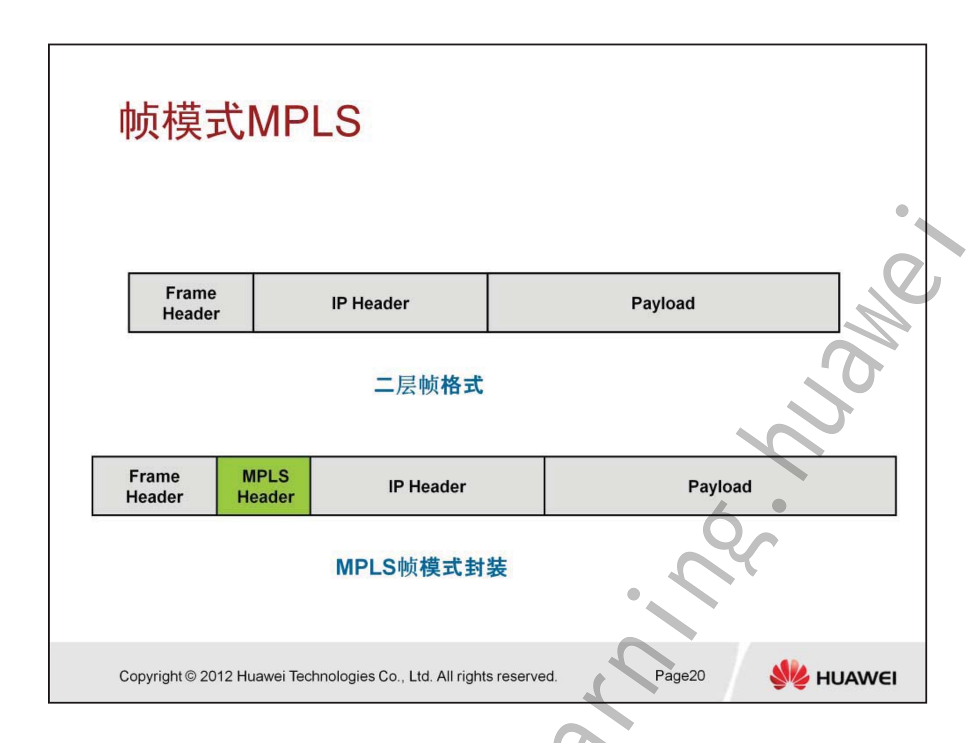
控制平面中路由协议模块（Routing Protocol）用来传递路由信息，生成路由信息表；标签分发协议模块（Label Distribution Protocol）用来完成标签信息的交换，建立标签转发路径。

数据平面包括IP转发表和标签转发表，当收到普通IP报文时（Incoming IP Packets），如果是普通IP转发，则查找IP路由表转发，如果需要标签转发，则按照标签转发表转发；当收到带有标签的报文时（Incoming Labeled Packets）时，如果需要按照标签转发，根据标签转发表转发，如果需要转发到IP网络，则去掉标签后根据IP转发表转发。

更多资料获取：http://learning.huawei.com/cn

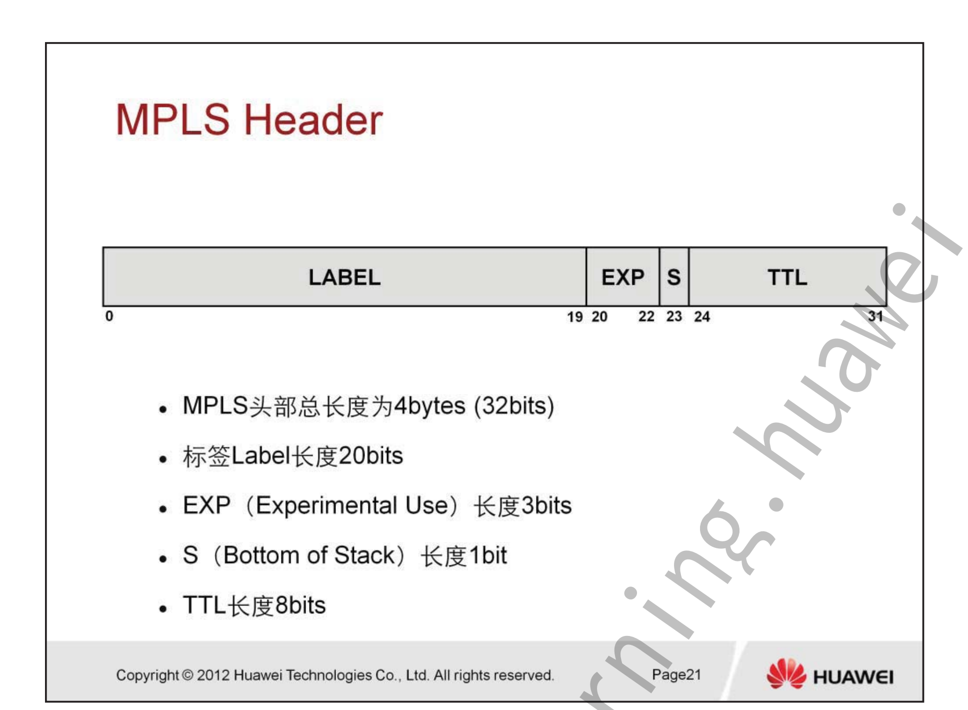


更多资料获取：http://learning.huawei.com/cn



MPLS有两种封装模式：帧模式和信元模式。帧模式封装是直接在报文的二层头部和三层头部之间增加一个MPLS标签头。以太网、PPP采用这种封装模式。

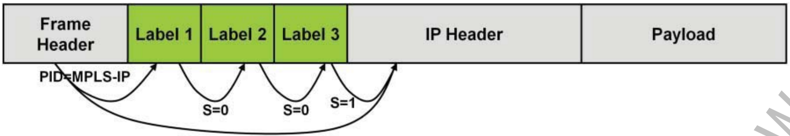
更多资料获取：http://learning.huawei.com/cn



MPLS Header长度为32bits，包括长度为20bits的标签，该标签用于报文转发；长度为3bits的EXP通常用来承载IP报文中的优先级；长度为1bit的栈底标志S用来表明是否是最后一个标签（MPLS标签可以多层嵌套）；长度为8bits的TTL，作用类似IP头部的TTL，用来防止报文环路等。

更多资料获取：http://learning.huawei.com/cn





PID 标识二层头部后面的报文类型

- Ethernet 0x0800 IPv4 0x8847 MPLS 单播报文 0x8848 MPLS多播报文
- PPP 0x8021 IPv4 0x8281 MPLS 单播报文 0x8283 MPLS多播报文

S 标识是否是栈底标签

标签嵌套应用

- MPLS VPN
- MPLS TE

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page22



二层报文头部中的协议字段PID可以用来标识该二层头部后面封装的是带标签的报文还是IP头，如Ethernet协议中PID=0x8847表示Ethernet头部后面封装的是MPLS 单播报文； PID=0x8848 表示 Ethernet头部后面封装的是MPLS多播报文； PID=0x0800 表示 Ethernet头部后面封装的是IPv4报文；PPP协议中，PID= 0x8281表示PPP头部后面封装的是MPLS 单播报文； PID= 0x8283表示PPP头部后面封装的是MPLS多播报文。

MPLS Header中的S字段可以用来表示其后面跟随的是另外一个标签还是三层的IP头。

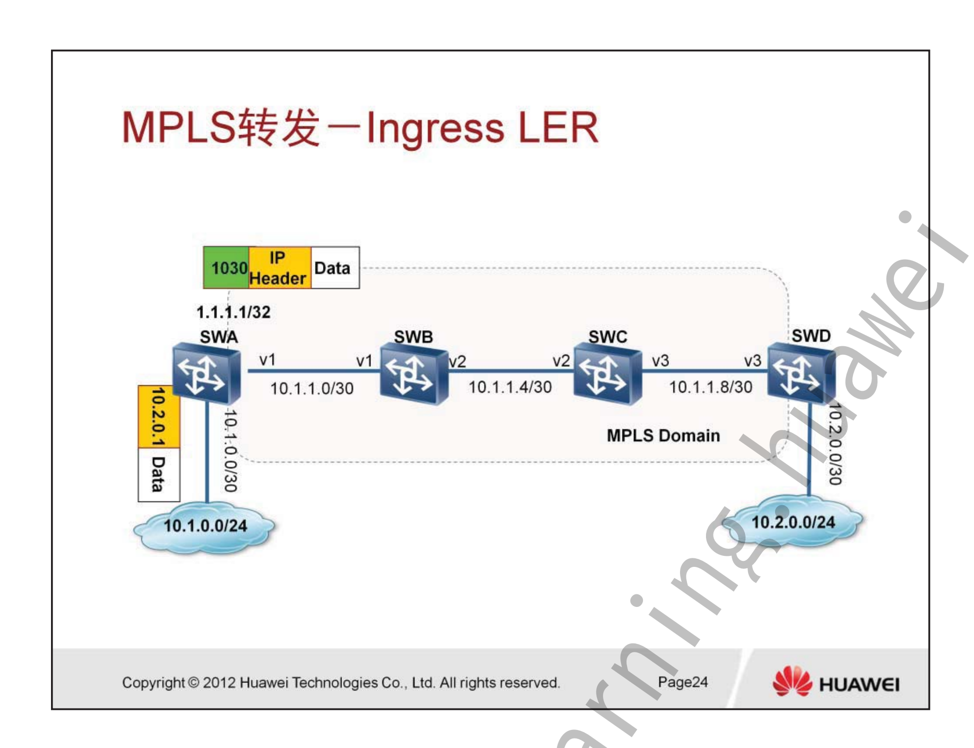
MPLS通常只为报文分配一个标签。但是在MPLS的高级应用会使用多层标签。如：MPLS VPN中会使用两个标签（复杂情况下，会用到三个标签），外层标签用于公网转发，内层标签用来标识报文属于哪个VPN；MPLS TE也会使用两个或多个标签，最外层标签标识TE隧道，内层标签表明报文的目的地。

注意：这里的Label1，Label2，Label3都指的是前一个胶片中的4个Bytes的MPLS头部，其中包含有20bits的标签信息。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



当一个IP报文进入MPLS域时，入口LER（SWA）会分析报文，根据该报文的特点（一般根据目的地址前缀分析）决定应该给该报文封装哪个标签以及应该从哪个接口转发给哪个下一跳。

更多资料获取：http://learning.huawei.com/cn

MPLS转发—Ingress LER

```
<SWA>display  mpls lsp include 10.2.0.0 24 verbose
---------------------------------------------------
              LSP Information: LDP LSP
---------------------------------------------------
 No                      :  1
 VrfIndex                :
 Fec                     :  10.2.0.0/24
 Nexthop                 :  10.1.1.2
 In-Label                :  NULL
 Out-Label               :  1030
 In-Interface            :  ----------
 Out-Interface           :  Vlanif1
 LspIndex                :  10249
 Token                   :  0x22005
 LsrType                 :  Ingress
 Outgoing token          :  0x0
 Label Operation         :  PUSH
 Mpls-Mtu                :  1500
 TimeStamp               :  822sec
```

- *FEC：Forwarding Equivalence Classes（转发等价类）*
- **NHLFE：Next Hop Label Forwarding Entry（下一跳标签转发表项）**

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page25

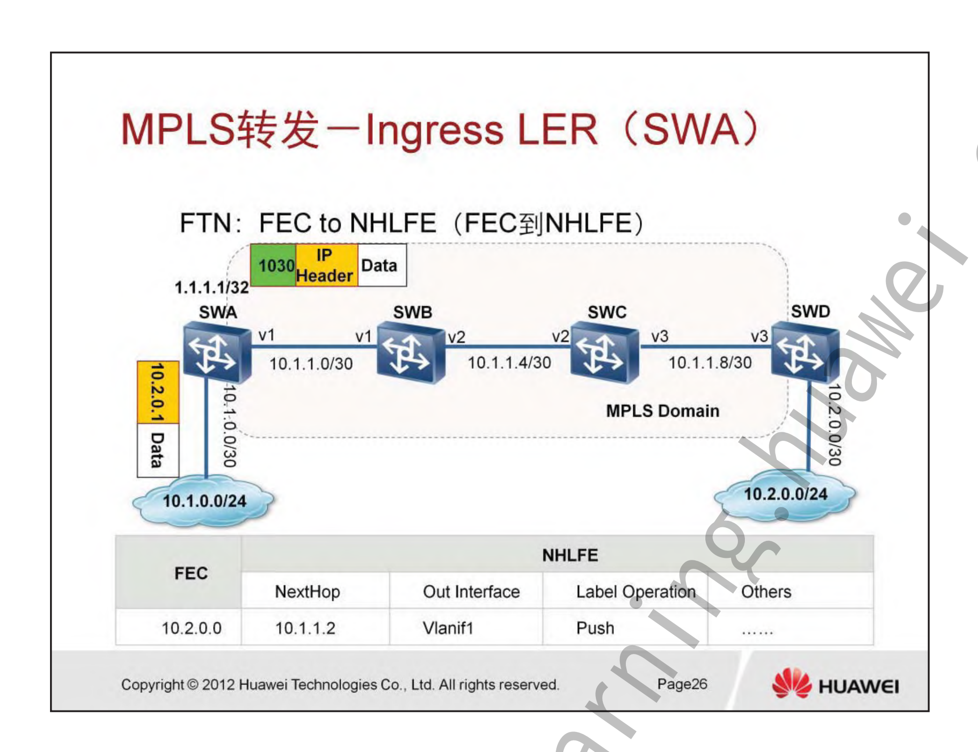


FEC（Forwarding Equivalence Class）是在转发过程中以等价的方式处理的一组数据分组，例如目的地址前缀相同的数据分组。通常对一个FEC分配唯一的标签。如本例中目的地址前缀为10.2.0.0/24的报文属于一个FEC，该FEC分到的标签为1030。

NHLFE（Next Hop Label Forwarding Entry）：进行标签转发时用到，NHLFE包含这样一些基本信息：1、报文的下一跳 2、如何进行标签操作（包括压入新的标签，弹出标签，用新的标签替换原有的标签等操作）。NHLFE还可能包含一些其他信息，如发送报文使用的链路层封装等。如本例中下一跳为10.1.1.2，标签操作为压入标签。



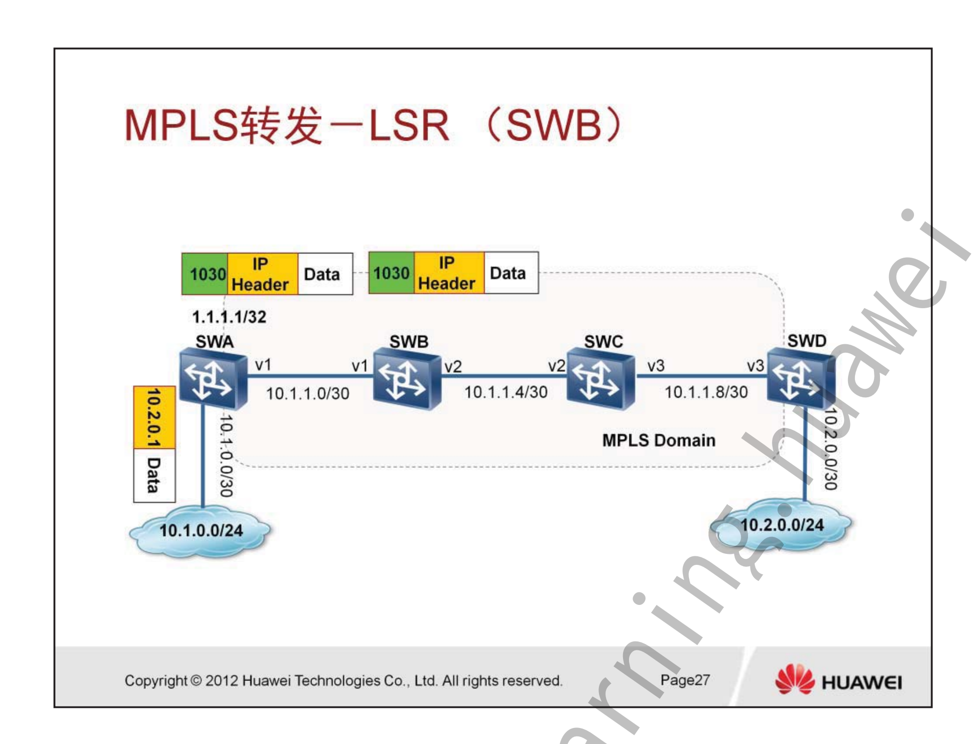
更多资料获取：http://learning.huawei.com/cn



FEC	NHLFE			
	NextHop	Out Interface	Label Operation	Others
10.2.0.0	10.1.1.2	Vlanif1	Push	……

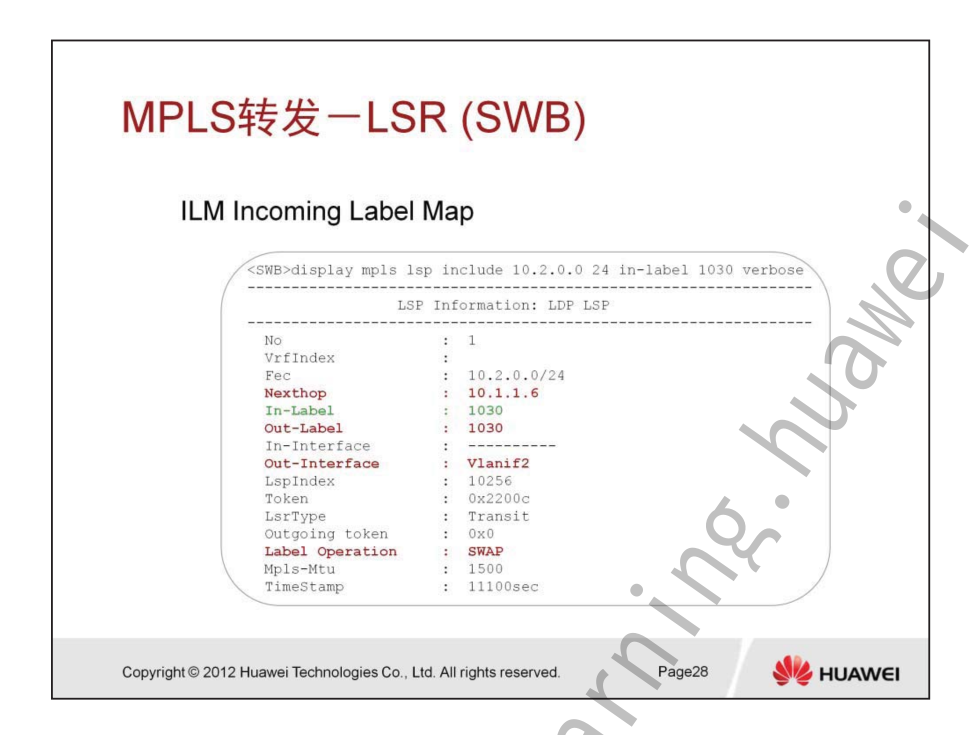
FEC代表了同一类报文，NHLFE包含了下一跳和标签操作等信息。只有将FEC和NHLFE关联起来，才能实现对于同一类报文进行特定的标签转发，FTN实现这个功能，FTN（FEC-to-NHLFE）将FEC映射到NHLFE。当LER将一个不带MPLS标签的IP报文转发给MPLS LSR时需要使用FTN。如果网络中存在等值路径，那么一个FEC可能会映射到多个NHLFE。

更多资料获取：http://learning.huawei.com/cn



SWB从SWA收到带有MPLS标签1030的报文，根据MPLS标签进行转发。

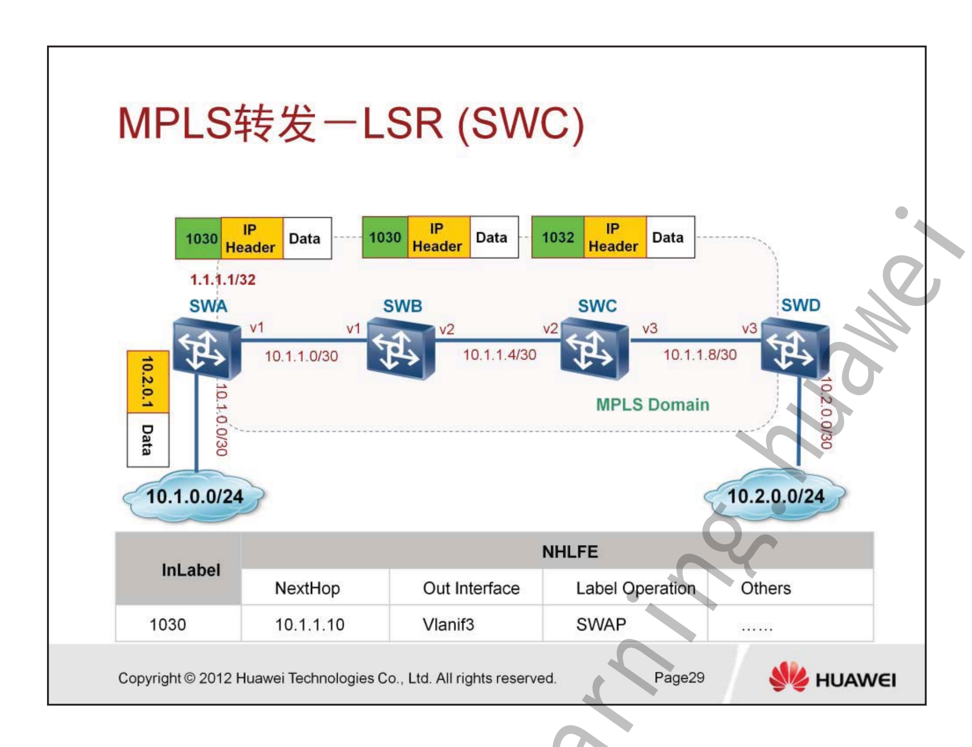
更多资料获取：http://learning.huawei.com/cn



本例中SWB收到带有标签1030的数据包，根据标签转发，找到下一跳为10.1.1.6，并用出标签替换入（SWAP）标签，继续转发。（本例情况特殊，出标签和入标签相同）。

ILM将每个入标签映射到NHLFE，当LSR转发带有标签的报文时使用ILM。同样，如果存在等值路径时一个入标签会映射多个NHLFE。

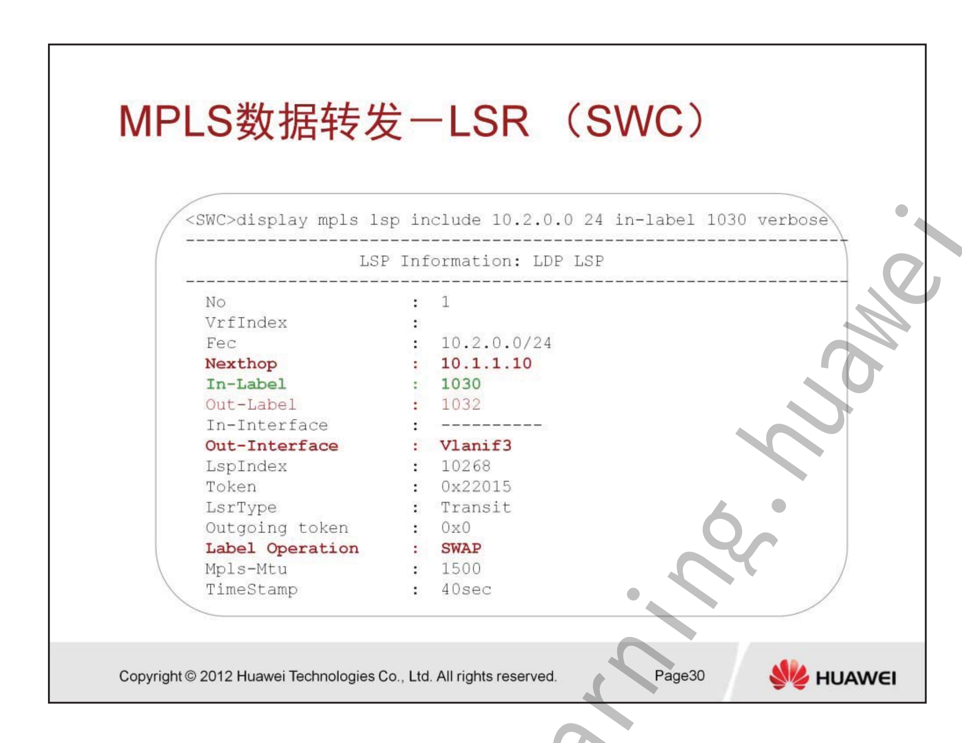
更多资料获取：http://learning.huawei.com/cn



InLabel	NHLFE			
	NextHop	Out Interface	Label Operation	Others
1030	10.1.1.10	Vlanif3	SWAP	……

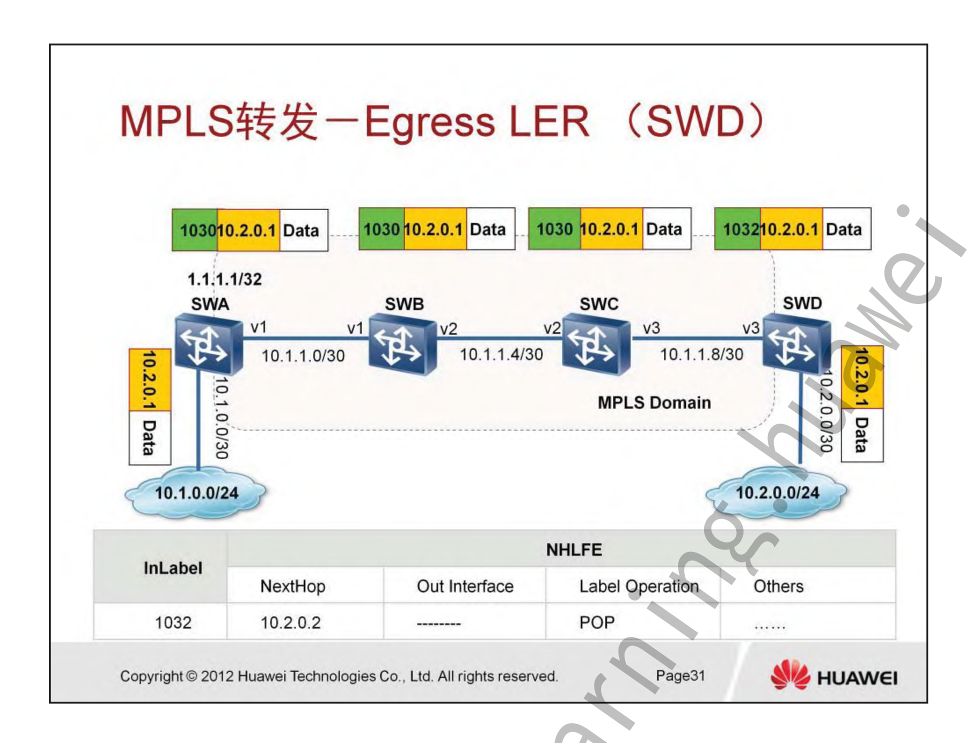
与SWB类似，SWC收到带有标签1030的报文后，根据标签进行转发，并用新出标签替换原来标签。

更多资料获取：http://learning.huawei.com/cn



本例中，SWC用出标签1032 替换（SWAP）入标签，然后从出接口Vlanif3转发，下一跳为10.1.1.10。

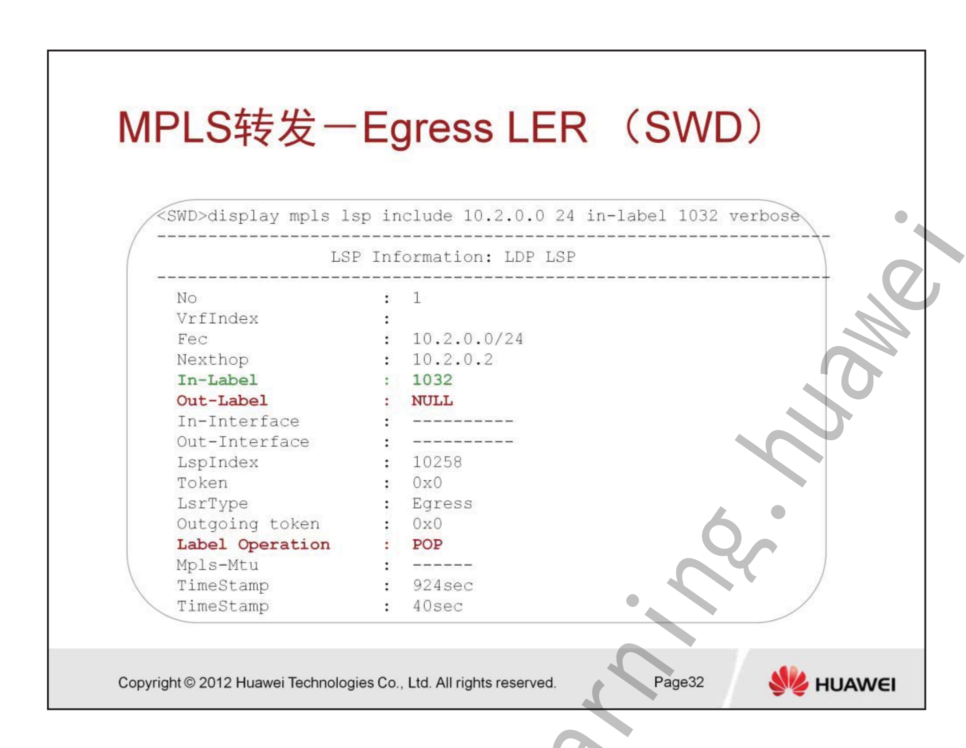
更多资料获取：http://learning.huawei.com/cn



InLabel	NHLFE			
	NextHop	Out Interface	Label Operation	Others
1032	10.2.0.2	--------	POP	……

Egress LSR SWD收到带有标签1032的报文后，去掉（POP）标签，查找IP路由表转发。

更多资料获取：http://learning.huawei.com/cn



本例中，SWD去掉标签1032，转发报文到下一跳10.2.0.2。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn

MPLS环路检测

IGP环路检测机制

TTL环路检测

- 帧模式的MPLS中使用TTL
- 信元模式的MPLS中无TTL

LDP环路检测机制

- 距离向量法
- 最大跳数法

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.　Page35

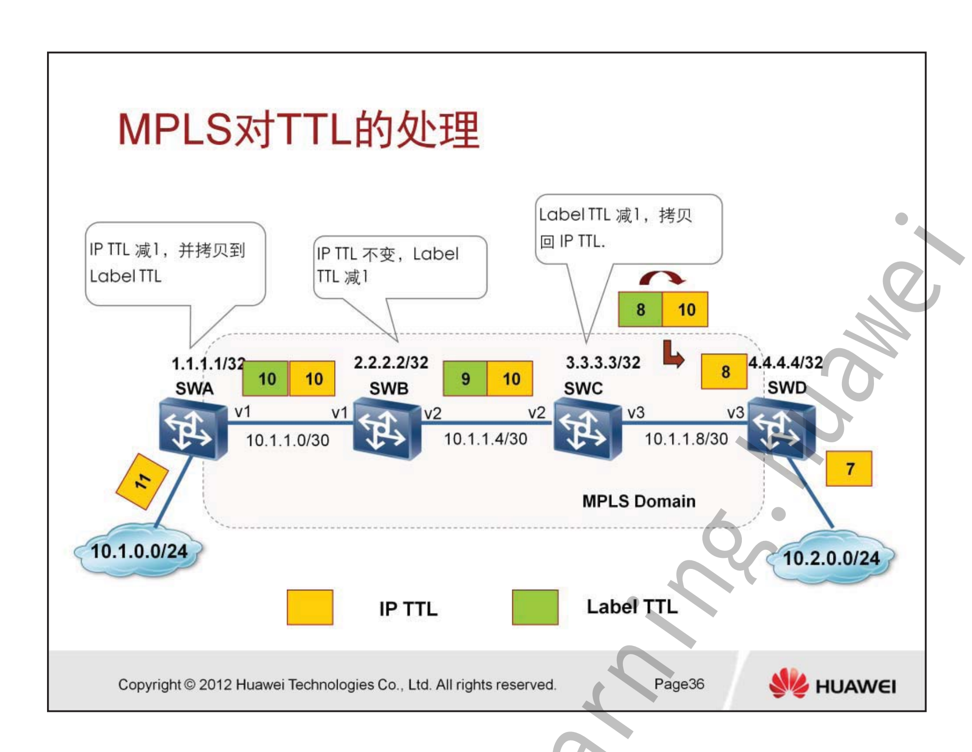


MPLS采用多种机制检测环路。

MPLS要依靠IGP建立LSP，IGP本身都有一些机制可以避免路由自环。单播IP报文的MPLS转发使用IGP产生的无环路的路径。

如果由于路由配置错误而产生了环路，在IP网络中可以使用TTL防止报文的无限循环转发，在MPLS网络中，也可以使用MPLS标签头部的TTL域来防止报文的无限循环转发。但是只有在帧模式的MPLS封装中才有TTL，在信元模式的MPLS封装中没有TTL，所以LDP又采用了一些特殊的机制检测环路。LDP环路检测方法有路径向量法和最大跳数法，这两种方法可以防止无限循环发送Label Request Message，可以通过配置选择是否使用这些方法。

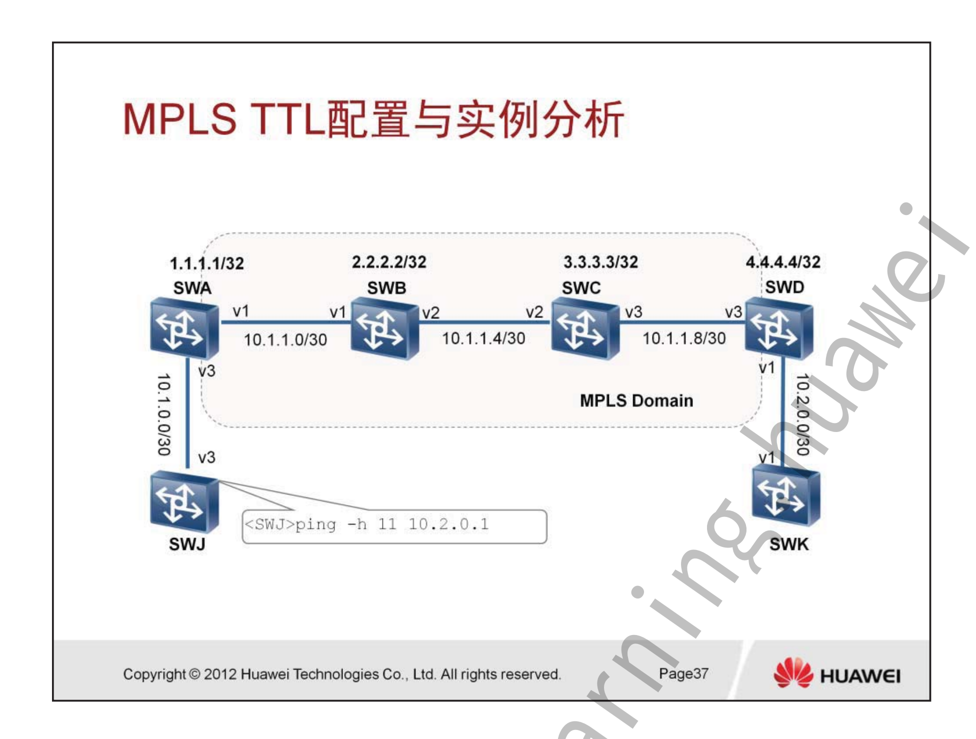
更多资料获取：http://learning.huawei.com/cn



MPLS对于TTL的处理有两种方式。一种是IP报文在进入MPLS网络的时候MPLS头部的TTL拷贝IP TTL值；另外一种是在入口LER将MPLS头部的TTL统一设置为255。

上图描述的是第一种情况。SWA从IP网络收到IP报文，IP TTL=11，SWA首先将IP TTL减1然后拷贝到MPLS的TTL域（IP TTL=10，Label TTL=10），SWB收到带有标签的报文，不处理IP TTL，只将MPLS的TTL减1（IP TTL=10，Label TTL=9），然后发送给SWC，SWC收到带有标签的报文，将MPLS的TTL减1（Label TTL=8），由于SWC为倒数第二跳，所以弹出标签，同时将MPLS的TTL拷贝到IP TTL中（IP TTL=8），SWD从SWC接收IP报文，按照普通IP转发将IP TTL减1后发送（TTL=7）。

更多资料获取：http://learning.huawei.com/cn



图中各台交换机之间正确运行路由协议，相互能够学习到各自的路由信息。SWA，SWB，SWC，SWD上运行启用MPLS和LDP并正常建立LDP Session分发标签建立LSP。

在缺省配置下，MPLS TTL会拷贝IP TTL值。

在SWJ上带TTL值ping SWK，观察SWA上的Debug信息。

在SWJ上tracert SWK，观察输出的信息。

更多资料获取：http://learning.huawei.com/cn

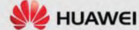
MPLS TTL配置与实例分析

```
<SWA>debug mpls packet
<SWA>debug ip packet acl 3000
<SWA>terminal monitor
<SWA>terminal debugging

*0.86297391 SWA IP/8/debug_case:
Receiving, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 2273, offset = 0, ttl = 11, protocol = 1,
checksum = 37572, s = 10.1.0.1, d = 10.2.0.1
prompt: Receiving IP packet from Serial3

*0.86297391 SWA IP/8/debug_case:
Sending, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 2273, offset = 0, ttl = 10, protocol = 1,
checksum = 37572, s = 10.1.0.1, d = 10.2.0.1
prompt: Sending the packet by lsp

*0.86297391 SWA MFW/8/MPLSFW PACKET:
PUSH Label=1030, EXP=0, TTL=10
Sending to V1, PktLen=88, Label(s)=1030, EXP=0, TTL=10
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page38 

命令<SWD>debug ip packet acl 3000 用来调试目的地址为10.2.0.1的IP报文。

命令<SWA>debug mpls packet用来调试MPLS报文。

从调试信息可以看出，SWA从接口V3收到IP TTL=11的IP报文，首先将IP TTL减1然后拷贝到MPLS的TTL域（IP TTL=10，Label TTL=10）。

更多资料获取：http://learning.huawei.com/cn

MPLS TTL配置与实例分析

```
<SWB>debug mpls packet
<SWB>debug ip packet acl 3000
<SWB>terminal monitor
<SWB>terminal debugging

*0.189653734 SWB MFW/8/MPLSFW PACKET:
Receiving from V1, PktLen=88, Label(s)=1030, EXP=0, TTL=10
SWAP Label=1029, EXP=0, TTL=9
Sending to V2, PktLen=88, Label(s)=1029, EXP=0, TTL=9
```

```
<SWC>debug mpls packet
<SWC>debug ip packet acl 3000
<SWC>terminal monitor
<SWC>terminal debugging

*0.189533719 SWC MFW/8/MPLSFW PACKET:
Receiving from V2, PktLen=88, Label(s)=1029, EXP=0, TTL=9
SWAP Label=3, TTL=8
Sending to V3, Dest=10.2.0.1, Nexthop=10.1.1.10
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

从调试信息可以看出，SWB从接口V1收到带标签的报文，MPLS TTL减1后TTL=9。SWC从接口V2收到带标签的报文，MPLS TTL减1后TTL=8并拷贝到IP TTL域将报文进行IP转发。

更多资料获取：http://learning.huawei.com/cn

MPLS TTL配置与实例分析

```
<SWD>debug mpls packet
<SWD>debug ip packet acl 3000
<SWD>terminal monitor
<SWD>terminal debugging
*0.64991297 SWD IP/8/debug_case:
Receiving, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 2273, offset = 0, ttl = 8, protocol = 1,
checksum = 38340, s = 10.1.0.1, d = 10.2.0.1
prompt: Receiving IP packet from Serial3
*0.64991297 SWD IP/8/debug_case:
Sending, interface = Serial1, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 2273, offset = 0, ttl = 7, protocol = 1,
checksum = 38596, s = 10.1.0.1, d = 10.2.0.1
prompt: Sending the packet from Serial3 at Serial1
```

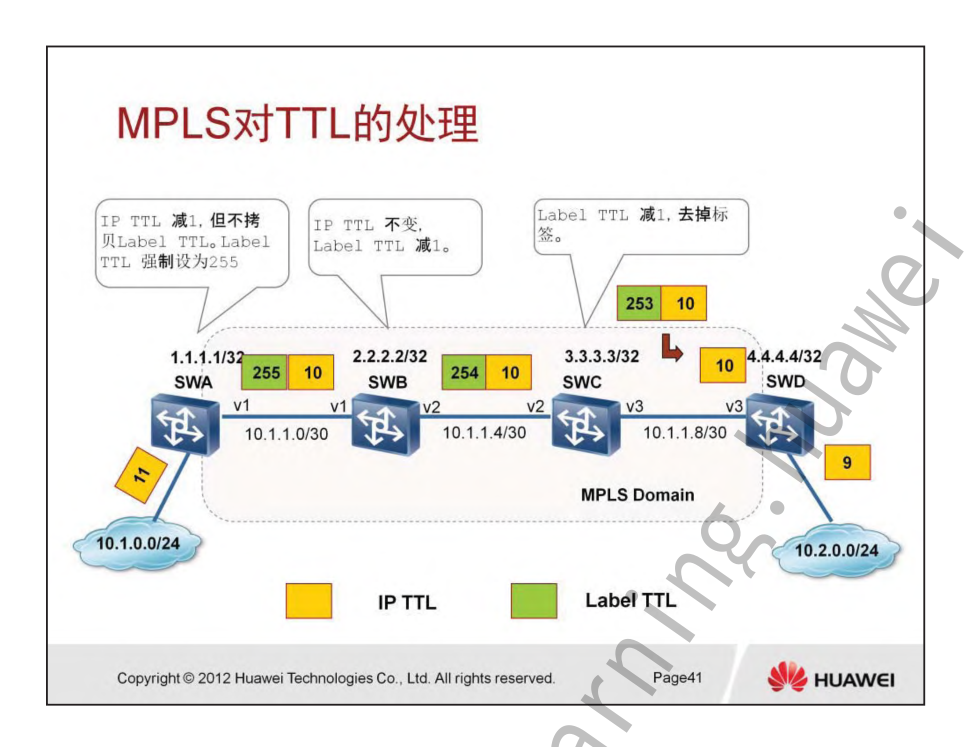
```
<SWJ>tracert 10.2.0.1
 traceroute to  10.2.0.1(10.2.0.1) 30 hops max,40 bytes packet
 1 10.1.0.2 31 ms  32 ms  1 ms
 2 10.1.1.2 62 ms  94 ms  62 ms
 3 10.1.1.6 94 ms  94 ms  94 ms
 4 10.1.1.10 125 ms  125 ms  125 ms
 5 10.2.0.1 156 ms  156 ms  156 ms
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page40 HUAWEI

SWD上从接口V3收到普通IP报文，TTL=8，按照普通IP包转发原理，SWD将TTL减1后TTL=7，然后从接口1转发。

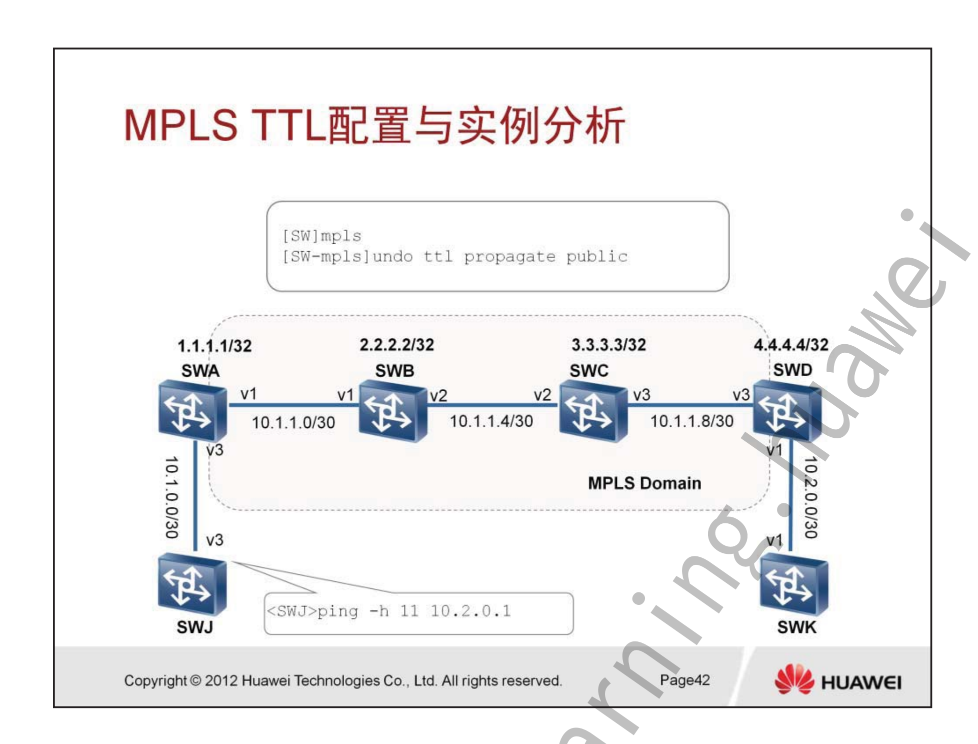
在SWJ上tracert 10.2.0.1，可以看到报文经过了每一台交换机，包括MPLS域内的所有LSR。

更多资料获取：http://learning.huawei.com/cn



这里描述第二种情况。SWA从IP网络收到IP报文，IP TTL=11，SWA将IP TTL减1，但并不将IP TTL拷贝到MPLS的TTL域，而是设置MPLS TTL为255（IP TTL=10，Label TTL=255），SWB收到带有标签的报文，不处理IP TTL，只将MPLS的TTL减1（IP TTL=11，Label TTL=254），然后发送给SWC，SWC收到带有标签的报文，将MPLS的TTL减1（Label TTL=253），由于SWC为倒数第二跳，所以弹出标签，将普通IP报文（TTL=10）发送给SWD，发送到SWD，SWD从SWC接收IP报文，按照普通IP转发将IP TTL减1后发送（TTL=9）。

更多资料获取：http://learning.huawei.com/cn



在SWA，SWB，SWC，SWD上配置MPLS不拷贝 IP TTL值。

在SWJ上带TTL ping SWK，观察SWA上的Debug信息。

在SWJ上tracert SWK，观察输出的信息。

配置解释：

[SW-mpls]undo ttl propagate public取消拷贝IP TTL值。

<SWJ>ping -h 11 10.2.0.1　发ping包到 10.2.0.1，TTL=11

更多资料获取：http://learning.huawei.com/cn

MPLS TTL配置与实例分析

```
<SWA>debug mpls packet
<SWA>debug ip packet acl 3000
<SWA>terminal monitor
<SWA>terminal debugging

*0.81886516 SWA IP/8/debug_case:
Receiving, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 1318, offset = 0, ttl = 11, protocol = 1,
checksum = 38527, s = 10.1.0.1, d = 10.2.0.1
prompt: Receiving IP packet from Serial3

*0.81886516 SWA IP/8/debug_case:
Sending, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 1318, offset = 0, ttl = 10, protocol = 1,
checksum = 38527, s = 10.1.0.1, d = 10.2.0.1
prompt: Sending the packet by lsp

*0.81886516 SWA MFW/8/MPLSFW PACKET:
PUSH Label=1030, EXP=0, TTL=255
Sending to V1, PktLen=88, Label(s)=1030, EXP=0, TTL=255
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page43 HUAWEI

可以看出SWA从IP网络收到IP TTL=11的IP报文，SWA将IP TTL减1，但并不将IP TTL拷贝到MPLS的TTL域，而是设置MPLS TTL为255（IP TTL=10，Label TTL=255）。

更多资料获取：http://learning.huawei.com/cn

MPLS TTL配置与实例分析

```
<SWD>debug mpls packet
<SWD>debug ip packet acl 3000
<SWD>terminal monitor
<SWD>terminal debugging

*0.99910344 SWD IP/8/debug_case:
Receiving, interface = Serial3, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 9625, offset = 0, ttl = 10, protocol = 1,
checksum = 30476, s = 10.1.0.1, d = 10.2.0.1
prompt: Receiving IP packet from Serial3

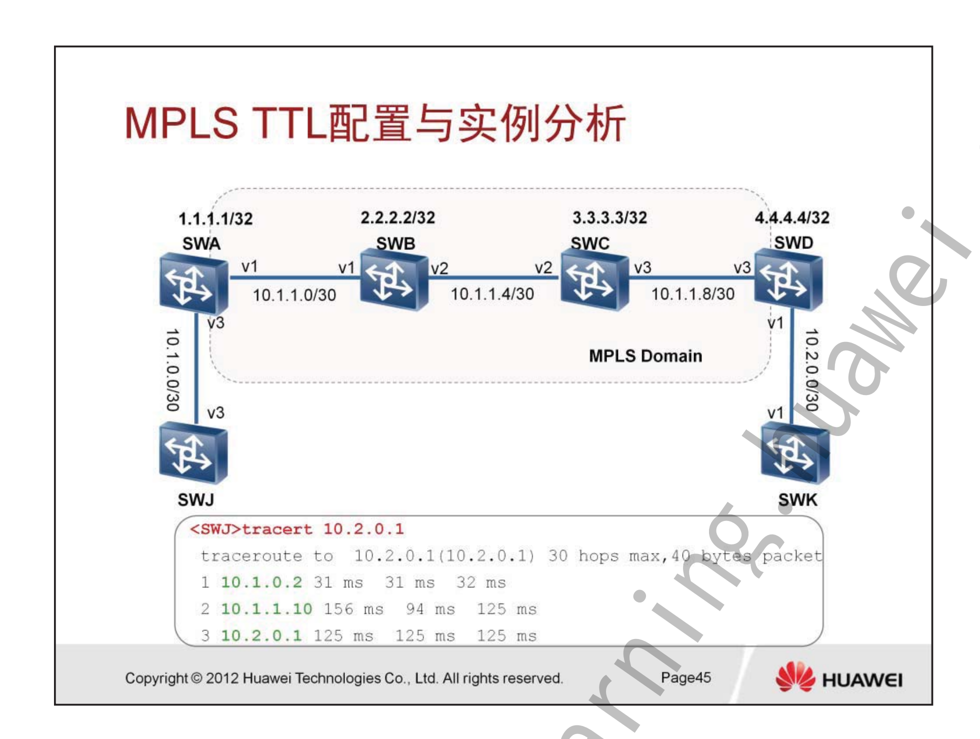
*0.99910344 SWD IP/8/debug_case:
Sending, interface = Serial1, version = 4, headlen = 20, tos = 0,
pktlen = 84, pktid = 9625, offset = 0, ttl = 9, protocol = 1,
checksum = 30732, s = 10.1.0.1, d = 10.2.0.1
prompt: Sending the packet from Serial3 at Serial1
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.
Page44



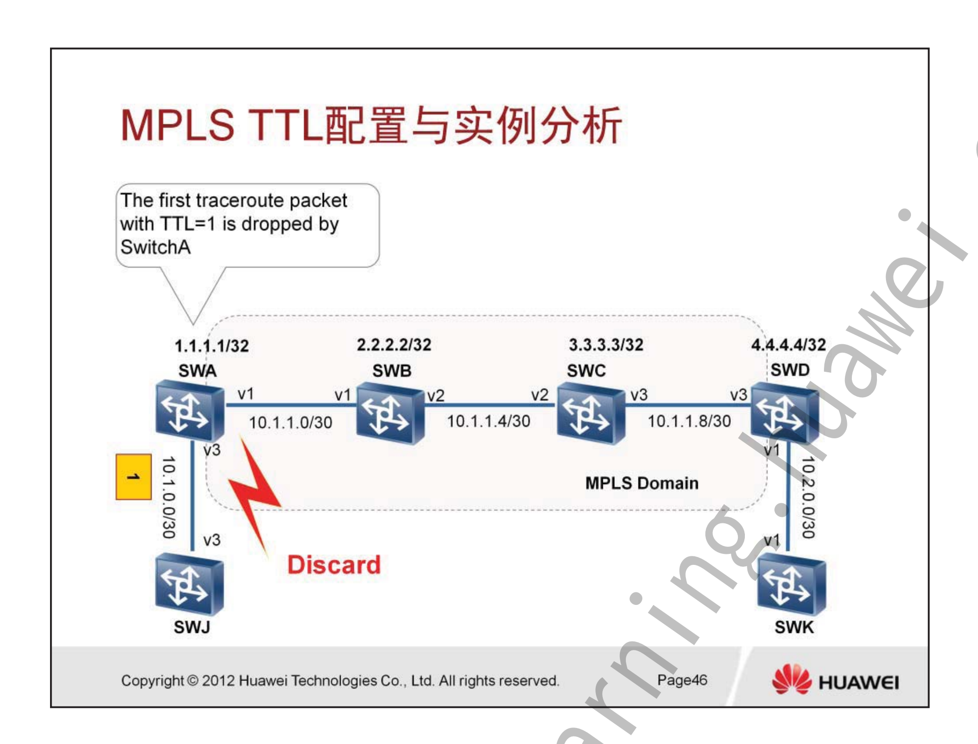
可以看出SWD从SWC收到IP TTL=10的IP报文（SWB和SWC根据标签转发，所以不修改报文的IP TTL值；SWC为倒数第二跳，去掉MPLS标签，将IP报文转发给SWD），SWD将IP TTL减1后转发到IP网络的SWK（IP TTL=9）。

更多资料获取：http://learning.huawei.com/cn



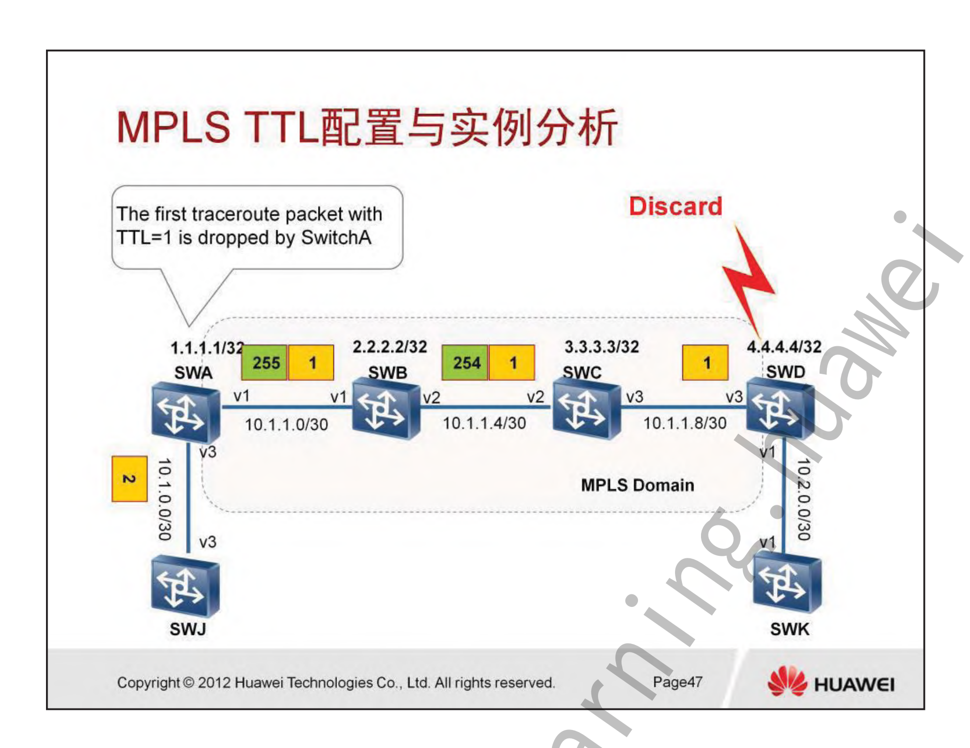
在SWJ上Tracert 10.2.0.1，可以看到在这种情况下，隐藏了MPLS域的LSR。

更多资料获取：http://learning.huawei.com/cn



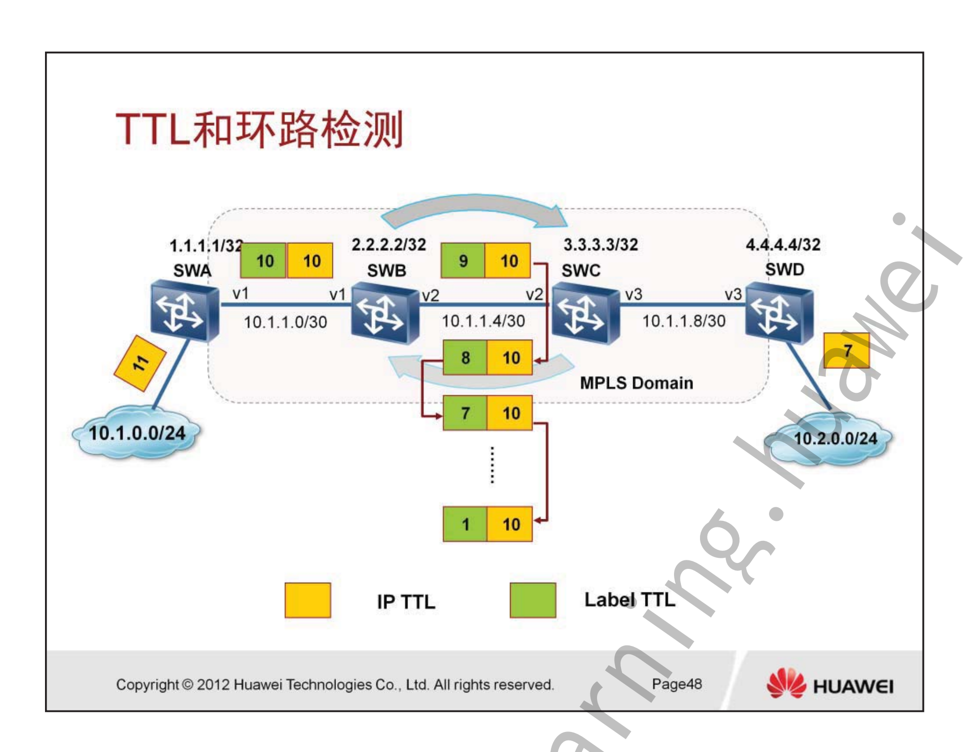
分析一下禁止了拷贝TTL的情况下从SWJ tracert SWK的情况。SWJ发出IP TTL=1的ICMP报文，SWA收到后减1为零，丢弃该报文并发送ICMP Reply给SWJ。

更多资料获取：http://learning.huawei.com/cn



接着SWJ重新发出TTL=2的报文，SWA收到该报文后，由于设置了MPLS不拷贝IP TTL，所以将IP TTL减1后保持不变，MPLS TTL设置为255，SWB只将MPLS TTL减1后继续发送，SWC不将MPLS TTL拷贝回IP TTL，而是弹出标签，将IP报文送到SWD，SWD将IP TTL减1后为0，丢弃该报文并返回ICMP Reply给SWJ。所以这种情况下，tracert只能看到LER而看不到经过的MPLS域的LSR。 TTL propagation可以用来进行故障定位，但要注意如果要undo ttl propagation，必须在MPLS域的所有交换机上使用该命令，以避免产生错误的结果。

更多资料获取：http://learning.huawei.com/cn

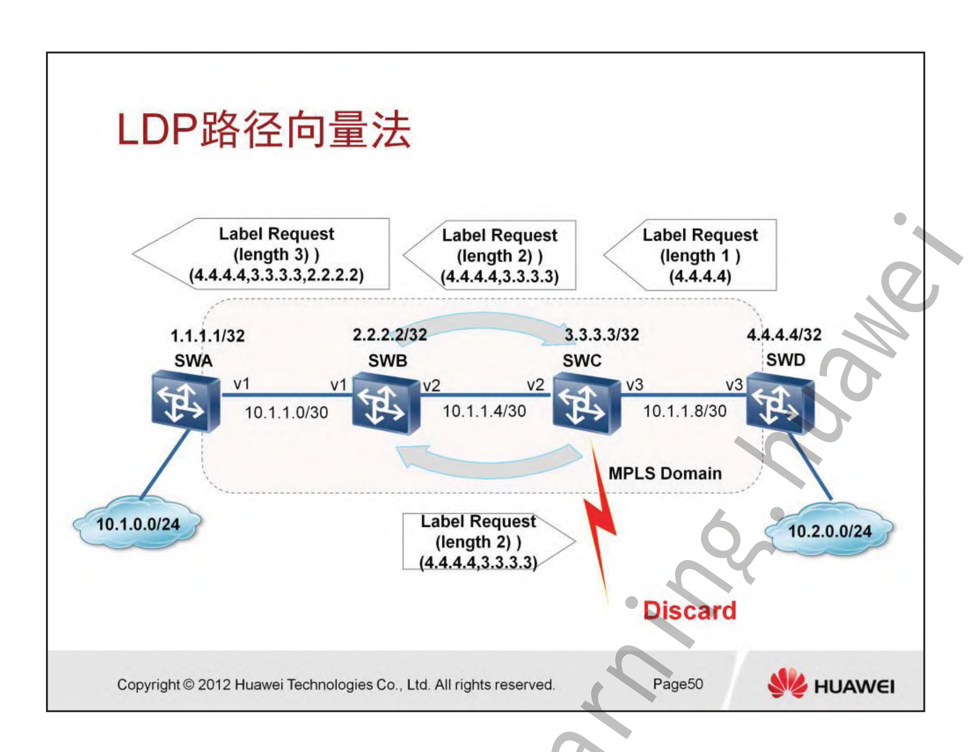


如果在SWB和SWC之间存在环路，带有标签的报文将会被SWB和SWC循环反复转发，每次MPLS TTL都会减1，直到 TTL=0，报文被丢弃。

更多资料获取：http://learning.huawei.com/cn



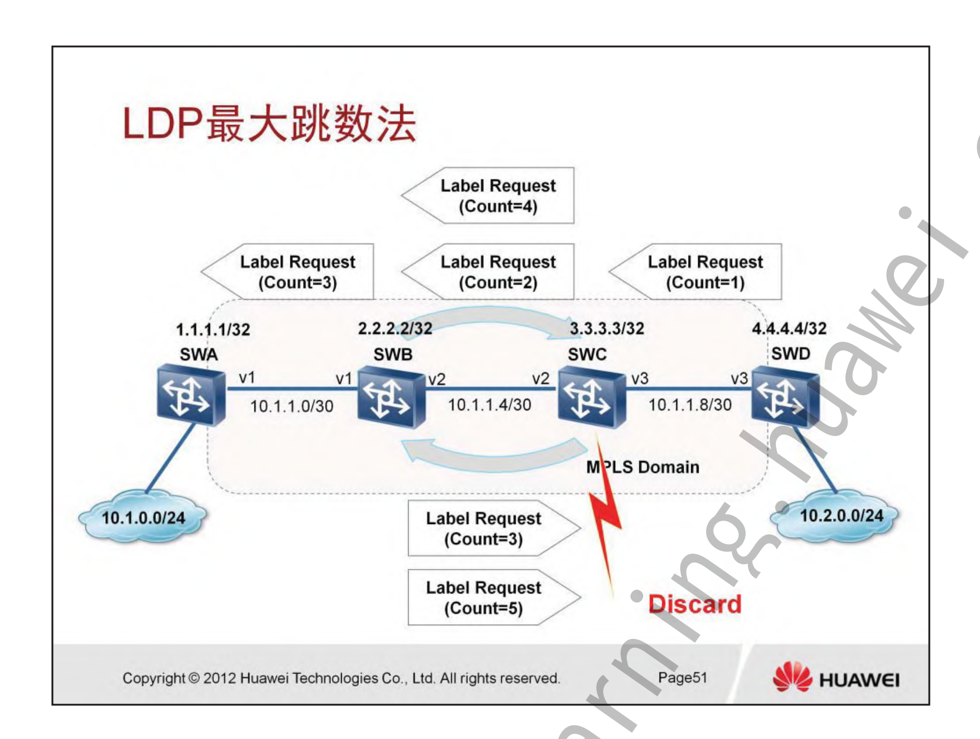
更多资料获取：http://learning.huawei.com/cn



LDP路径向量法和最大跳数法分别通过两类TLV实现：Path Vector TLV和 Hop Count TLV 。如果配置使用这两类方法检测环路，那么在标签请求消息（Label Request Message）和标签映射消息（Label Mapping Message）中都会携带这两种TLV。

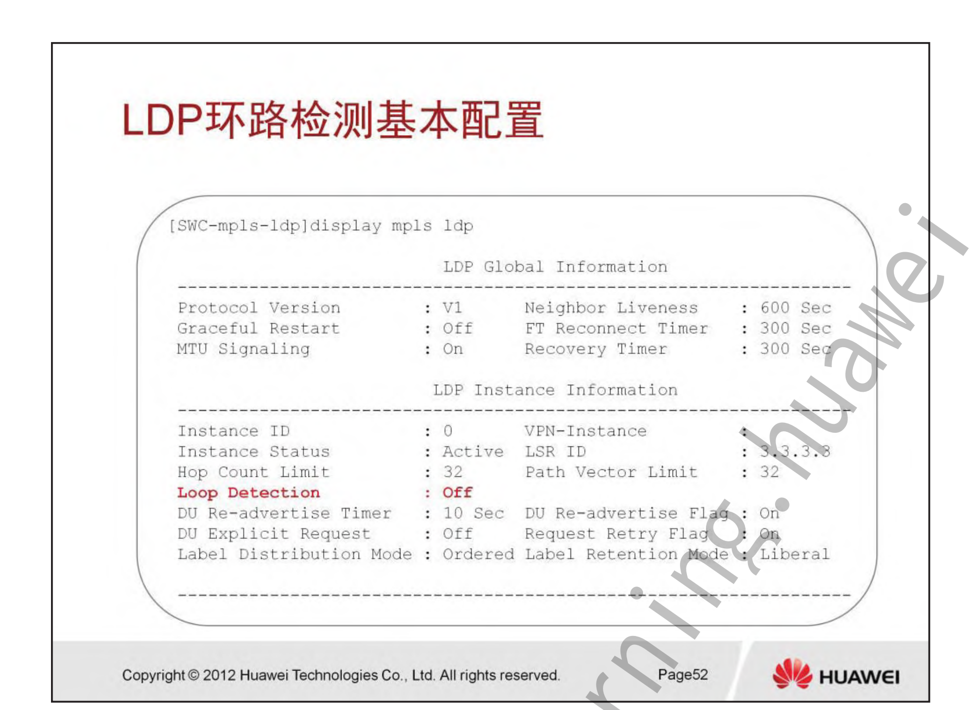
首先介绍距离向量法。每个LSR在发送标签请求消息（标签映射消息）中，包含一个Path Vector　TLV，并且入口（出口）LSR产生的路径长度值为1，并且把自己的LSR ID加到TLV的列表中，标签请求消息（标签映射消息）每经过一跳长度值加1；接收端LSR如果收到的标签请求消息（标签映射消息），发现长度值达到预先设定的最大值或者发现LSR ID列表中有自己的LSR ID，认为发现环路，发出通知消息，拒绝LSP的建立。

更多资料获取：http://learning.huawei.com/cn



使用最大跳数法时，每个LSR在发送标签请求消息（标签映射消息）中，包含一个hop count TLV，并且入口（出口）LSR产生的初始跳数值为1，标签请求消息（标签映射消息）每经过一跳跳数值加1；接收端LSR如果在收到的标签请求消息中（标签映射消息），发现跳数值达到预先设定的最大值，认为发现环路，发出通知消息，拒绝LSP的建立。

更多资料获取：http://learning.huawei.com/cn



缺省配置为不进行LDP环路检测。

更多资料获取：http://learning.huawei.com/cn

LDP环路检测基本配置

```
[SWC-mpls-ldp]
[SWC-mpls-ldp]loop-detect
Warning: Loop-Detection cannot be configured after enabling LDP
on an interface
[SWC-mpls-ldp]quit
[SWC]undo mpls ldp
[SWC]mpls ldp
[SWC-mpls-ldp]loop-detect
[SWC-mpls-ldp]hops-count ?
  INTEGER<1-32>  Value of the maximum Hop-Count
[SWC-mpls-ldp]path-vectors ?
  INTEGER<1-32>  Value of the Path-Vector limit
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page53 HUAWEI

如果对MPLS域进行环路检测，则必须在所有节点上都配置环路检测，并且需要在所有接口使能LDP之前进行配置。但在建立LDP会话时，并不要求双方的环路检测配置一致。

更多资料获取：http://learning.huawei.com/cn



问 题

MPLS转发是根据什么完成数据转发的?

MPLS常见应用有哪些?

MPLS封装有哪些方式，各自应用范围是什么?

MPLS有哪些环路检测方法?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page54 

答案:

MPLS转发是根据什么进行数据转发的?

- MPLS是根据标签进行数据转发的。

MPLS常见应用有哪些?

- MPLS VPN，MPLS QoS，MPLS TE。

MPLS封装有哪些方式，各自应用范围是什么?

- 帧模式和信元模式。Ethernet和PPP使用帧模式封装，ATM使用信元模式封装。

LDP邻居发现机制有哪两种，分别有什么区别?

- 基本发现机制和扩展发现机制，基本发现机制用来发现同一链路上的邻居，扩展发现机制用来发现非同一链路上的邻居。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



前 言

本课程介绍了LDP标签空间、标签分发协议、LDP邻居协议工作原理以及在VPN、QoS和流量工程方面的应用。

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page1



更多资料获取：http://learning.huawei.com/cn



培训目标

学完本课程后，您应该能：

- 描述LDP邻居发现机制
- 描述LDP会话建立过程
- 掌握LDP标签管理

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page2

HUAWEI

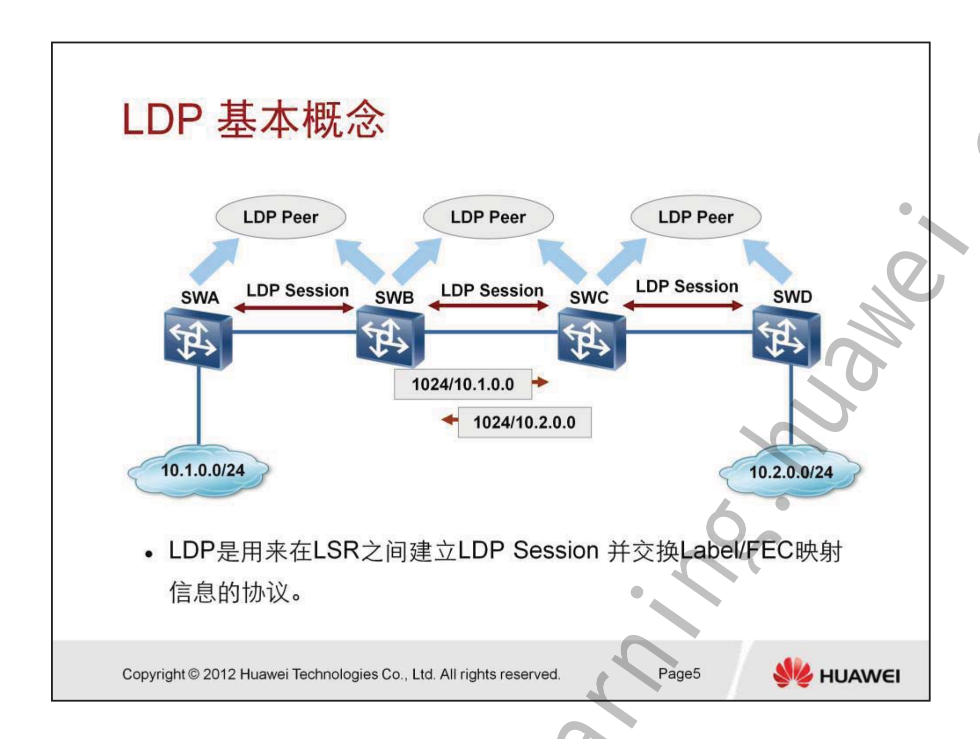
更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



MPLS需要使用标签分发协议完成标签的分配控制和保持，目前有很多种标签分发协议，LDP（Label Distribution Protocol）为其中之一，LSR之间可以使用LDP协议来交换标签信息。

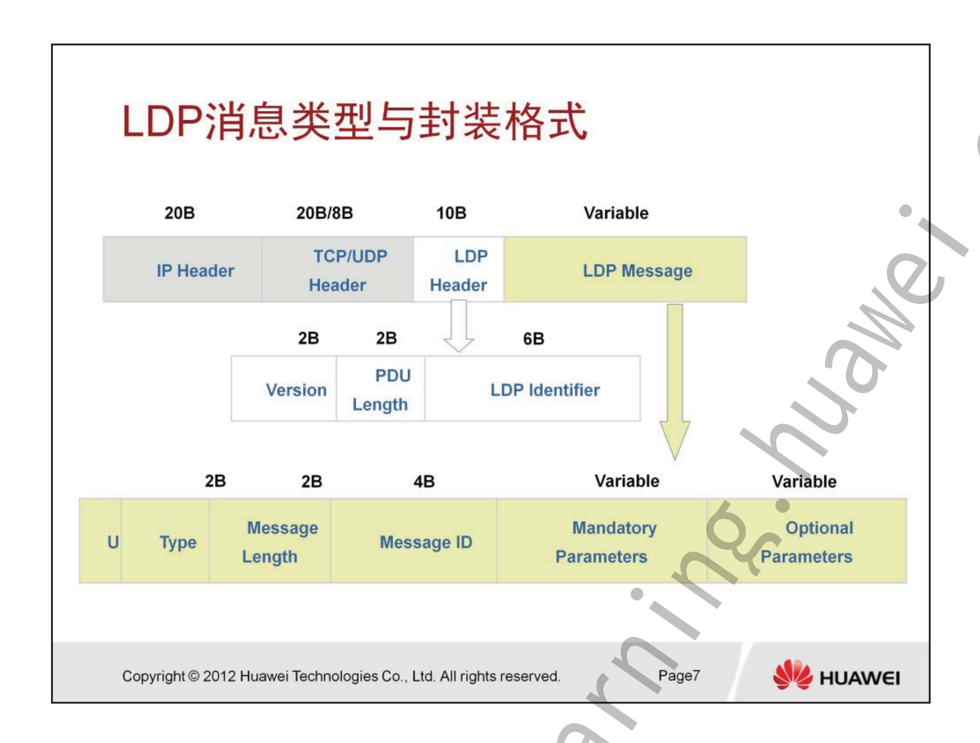
上图中，SWA，SWB，SWC，SWD配置为LSR，运行了LDP协议的两台LSR之间建立LDP Session并交换Label/FEC映射信息，建立了LDP Session的两台交换机称为LDP Peers。

更多资料获取：http://learning.huawei.com/cn

运行LDP协议的LSR之间通过交换LDP消息来发现邻居、建立和维护LDP Session并管理标签。 LDP消息承载在UDP或TCP之上，端口号为646 。这里简单介绍LDP常用的一些消息和各个消息的主要功能。按照消息的功能，LDP消息一共可以分为四大类型：Discovery Message，Session Message，Advertisement Message和Notification Message。Discovery message用来宣告和维护网络中一个LSR的存在；Session message用来建立、维护和终止LDP Peers之间的LDP Session；Advertisement message:用来生成、改变和删除FEC的标签映射；Notification message用来宣告告警和错误信息。

Discovery Message用来发现邻居，承载在UDP报文上。LDP要求可靠而有序地传递消息，所以LDP使用TCP建立Session，Session Message，Advertisement Message，Notification Message等消息都基于TCP传递。

更多资料获取 http://learning.huawei.com/cn



LDP PDU包括LDP Header和LDP Message两部分。

LDP Header长度为10Bytes，包括Version，PDU Length和LDP Identifier三部分。其中Version占用2Bytes，表示LDP版本号，当前版本号为1。PDU Length长度为2Bytes，以字节为单位表示除了Version和PDU Length以外的其他部分的总长度。LDP Identifier长度6Bytes，其中前4Bytes用来唯一标识一个LSR，后2Bytes用来表示LSR的标签空间，LSR的标签空间将当U=1时在“Part 4 LDP标签管理”中详细介绍。

LDP Message包含五个部分。其中U占用1个bit，为Unknown Message bit。当LSR收到一个无法识别的消息时，该消息的U=0时，LSR会返回给该消息的生成者一个通告，，忽略该无法识别的消息，不发送通告给生成者。Message Length占用2个bytes，以字节为单位表示Message ID、Mandatory Parameters和Optional Parameters的总长度。Message ID占用32个bit，用来标识一个消息。 Mandatory Parameters和Optional Parameters分别为可变长的该消息的必须的参数和可选的参数。 Message Type表示具体的消息类型，目前，LDP定义的常用的消息有Notification，Hello，Initialization，KeepAlive，Address，Address Withdraw，Label Mapping，Label Request，Label Abort Request，Label Withdraw，Label Release。

更多资料获取：http://learning.huawei.com/cn



消息类型	作用
Hello	LDP发现机制中宣告本LSR并发现邻居
Initialization	在LDP Session建立过程中协商参数
KeepAlive	监控LDP Session的TCP连接的完整性
Address	宣告接口地址
Address Withdraw	撤消接口地址
Label Mapping	宣告FEC/Label映射信息
Label Request	请求FEC的标签映射
Label Abort Request	终止未完成的Label Request Message
Label Withdraw	撤消FEC/Label映射
Label Release	释放标签
Notification	通知LDP Peer错误信息

各个消息的主要作用如下:

Notification Message用来通知LDP Peer错误信息或者其他一些辅助信息如LDP Session状态等。

Hello Message主要用来在LDP发现机制中宣告本LSR并发现邻居。

Initialization Message用来在LDP Session建立过程中协商参数。

KeepAlive Message用来监控LDP Session的TCP连接的完整性。

Address Message用来宣告接口地址。

Address Withdraw Message用来撤消接口地址。

Label Mapping Message用来宣告FEC/Label映射信息。

Label Request Message 用来向LDP Peer请求FEC的标签映射。

Label Abort Request Message用来终止未完成的Label Request Message。

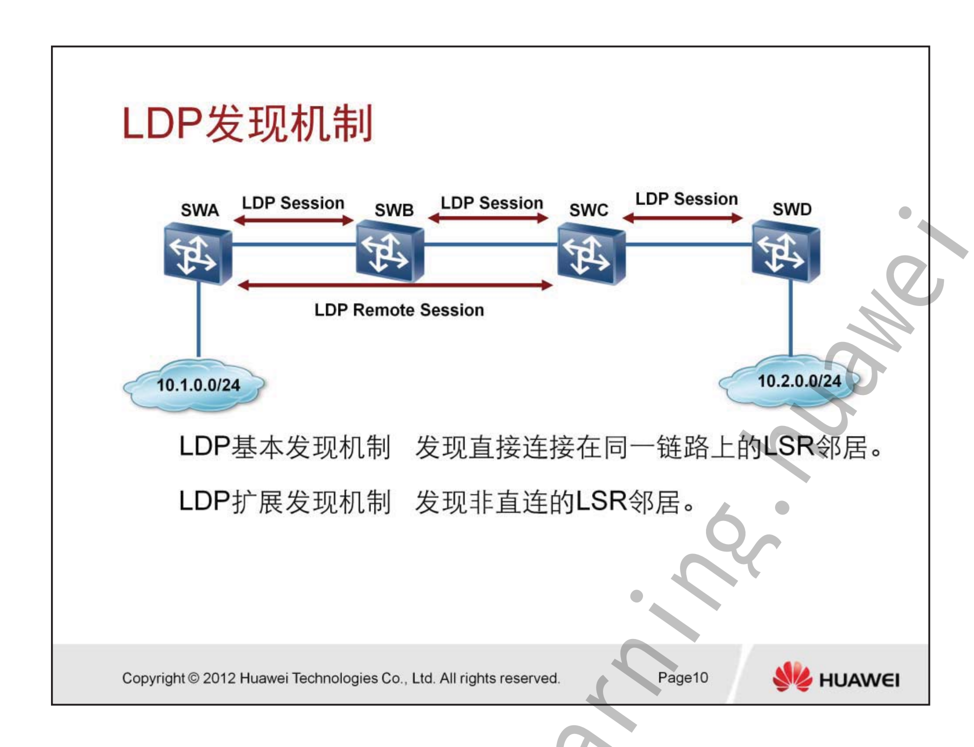
Label Withdraw Message用来撤消FEC/Label映射。LSR通过发送Label Withdraw Message告诉对等体该对等体不可以继续使用自己以前通告给他的标签。

Label Release Message用来释放标签。当一个LSR不再需要以前从LDP Peer收到的标签时，就发送一个Label Release Message给该LDP Peer。

更多资料获取：http://learning.huawei.com/cn

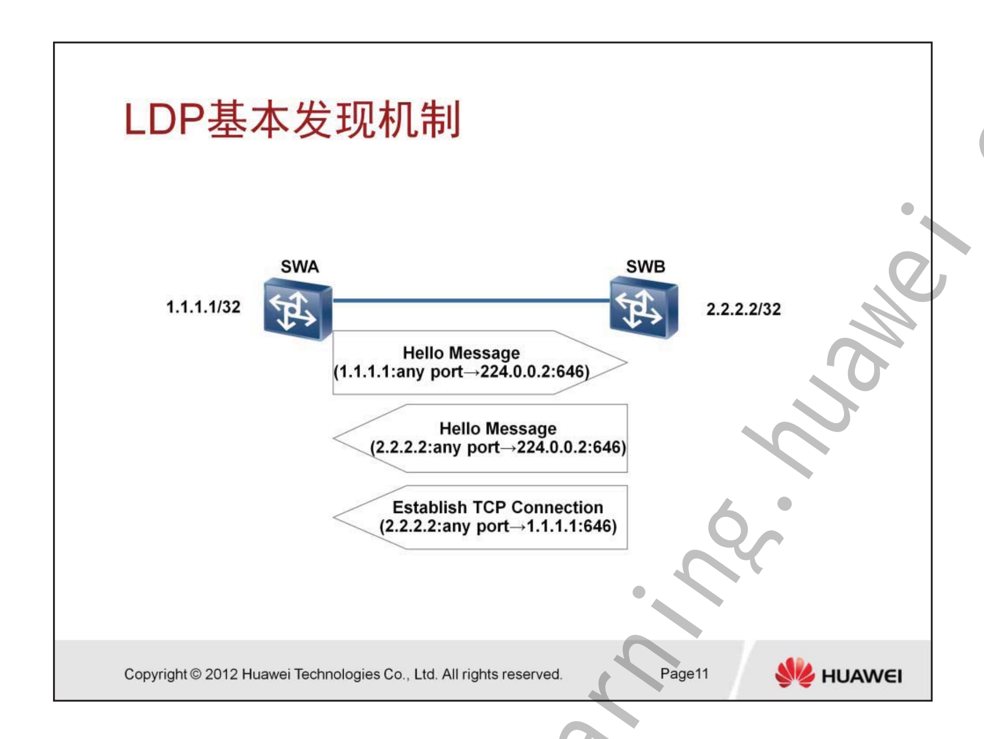


更多资料获取：http://learning.huawei.com/cn



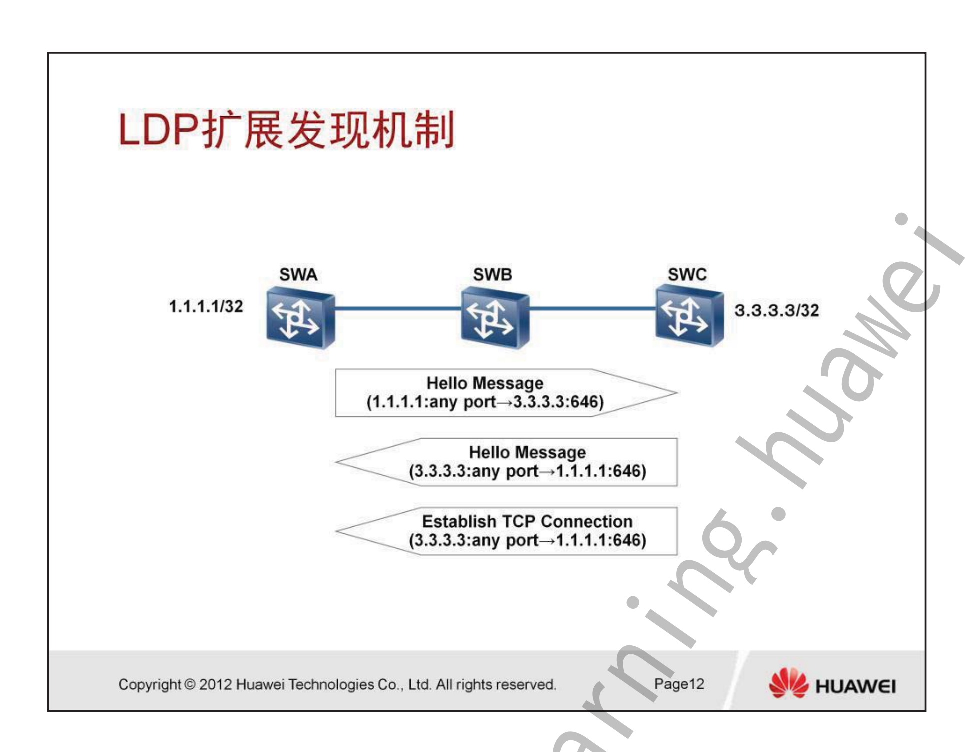
LSR通过LDP发现机制发现LDP Peers。LDP发现机制包括LDP基本发现机制和LDP扩展发现机制。LDP基本发现机制可以自动发现直连在同一条链路上的LDP Peers，所以这种情况下不需要明确指明LDP Peer；LDP扩展发现机制能够发现非直连的LDP Peers。

更多资料获取：http://learning.huawei.com/cn



LDP的Discovery message用于邻居发现，他提供了这样一个机制：LSR通过周期性地发送Hello Message表明自己的存在。这个消息是封装在UDP报文中的，目的端口号为646。在LDP基本发现机制中，该消息的目的IP地址为组播IP地址224.0.0.2，即该消息发给该网段上所有的交换机（如图中的SWA和SWB分别周期性地发送Hello Message给224.0.0.2）。Hello Message中携带了LDP Identifier信息以便告诉对方自己使用的标签空间。然后IP地址大的LSR作为主动方发起TCP连接。TCP连接建立之后，LSR会继续发送Hello Message以便发现新的邻居或者检测错误。

更多资料获取：http://learning.huawei.com/cn

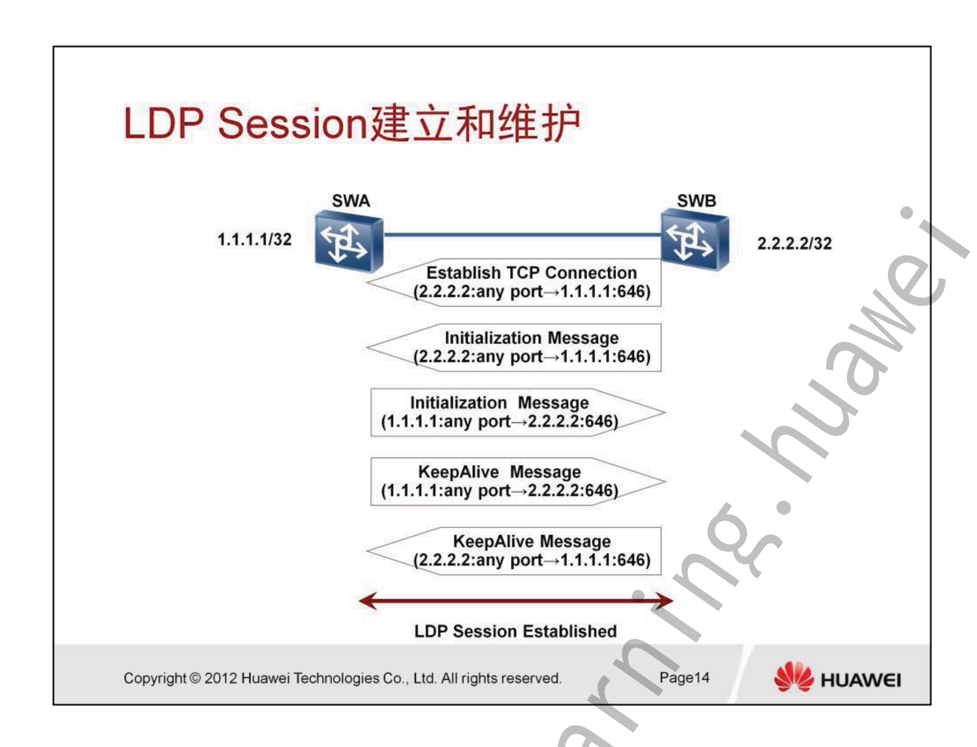


与LDP基本发现机制不同的是，LDP扩展发现机制是运行LDP协议的LSR（如SWA）周期性地发送Hello Message给特定的目的IP，所以需要通过配置指定建立Session的LDP Peer，另外一个LSR（如SWB）将决定是否要回应该报文，如果要回应，则通过发送Hello Message给特定的LSR（SWA）。

更多资料获取：http://learning.huawei.com/cn



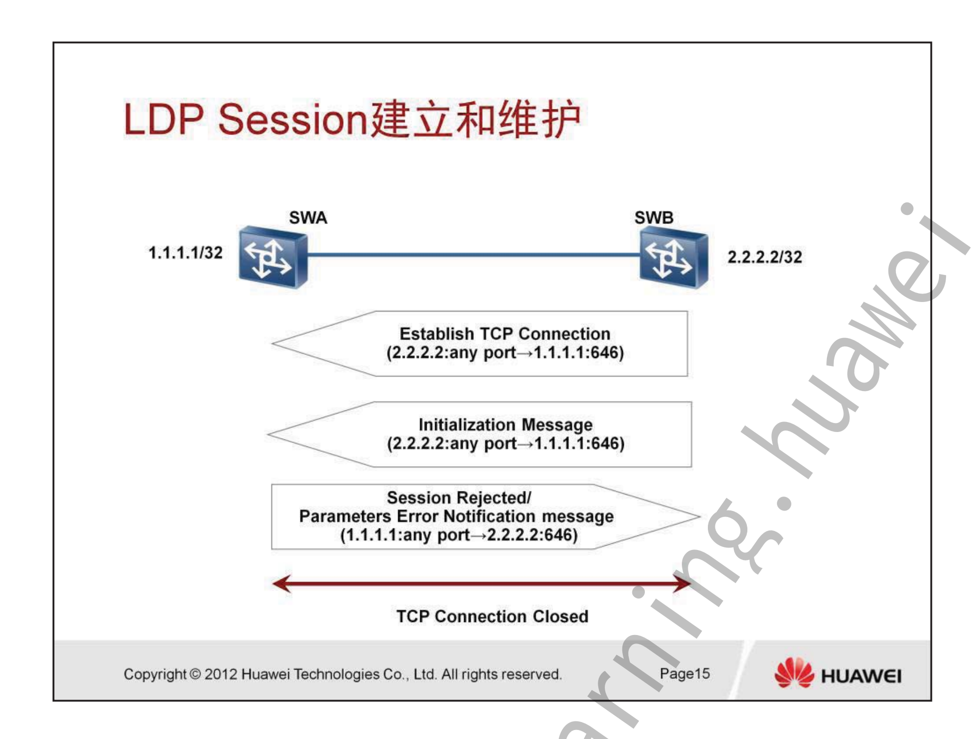
更多资料获取：http://learning.huawei.com/cn



建立TCP连接之前，两个LSR（如图中SWA和SWB）首先会决定使用哪个地址建立TCP连接以及谁是主动方谁是被动方，然后由主动方发起连接。如果Hello Message中携带了Transport Address，该Transport Address用于建立TCP连接，如果Hello Message中没有携带Transport Address，则该Hello Message的源IP地址用于建立TCP连接。两个LSR都从对端发来的Hello Message中获得对端用于建立TCP连接的地址，然后比较两个地址的大小，地址大的作为主动方发起TCP连接，图中SWB作为主动方发起TCP连接。

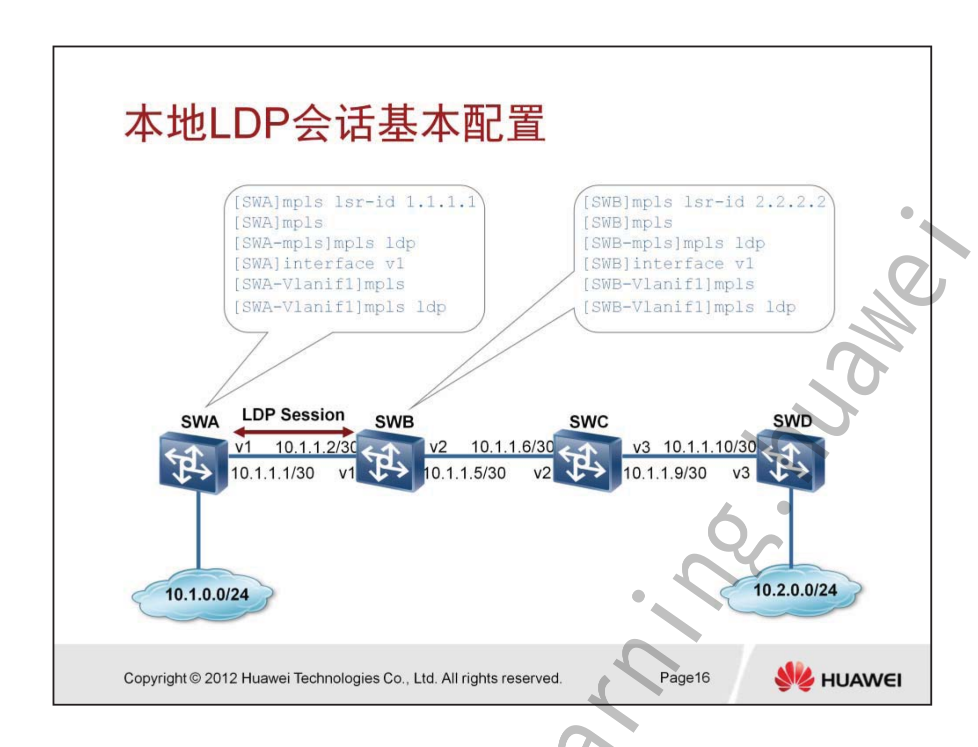
TCP连接建立之后，由主动方（SWB）发出Initialization Message携带会话协商参数，如LDP协议号，标签分发方式等等，被动方（SWA）检查参数能够接受，如果接受则发送Initialization Message并携带自己希望使用的协商参数，并随后发KeepAlive Message。直到双方都收到对端的KeepAlive Message后，会话建立。两个LSR就会成为LDP Peers并交换Advertisement Message。

更多资料获取：http://learning.huawei.com/cn



如果被动方（SWA）不接受协商参数，则发送Error Notification Message给对方取消连接。

更多资料获取：http://learning.huawei.com/cn



图中，SWA 的Loopback1地址为1.1.1.1/32，SWB 的Loopback1地址为2.2.2.2/32，SWC 的Loopback1地址为3.3.3.3/32，SWD 的Loopback1地址为4.4.4.4/32。各交换机之间的互连地址如图所示。

配置SWA和SWB之间正确建立LDP Session，以便分发标签。

配置基本思路：

1、首先需要配置SWA和SWB的mpls lsr-id，用于建立和维护LDP Session。VRP没有缺省的LSR ID，必须手工配置； lsr-id使用IPv4地址格式，在MPLS域内唯一；通常使用Loopback接口的IPv4地址作为LSR ID。

2、lsr-id是用来建立LDP Session，缺省情况下使用lsr-id建立TCP连接。所以交换机上要配置路由协议使得SWA和SWB的lsr-id之间可达。

3、全局模式下，使能MPLS和MPLS LDP。

4、相应的接口下，使能MPLS和MPLS LDP。

配置说明：

[SWA]mpls lsr-id 1.1.1.1

配置lsr-id。

[SWA]mpls

更多资料获取：http://learning.huawei.com/cn

全局模式下使能mpls

[SWA-mpls]mpls ldp

MPLS模式下或全局模式下使能ldp。

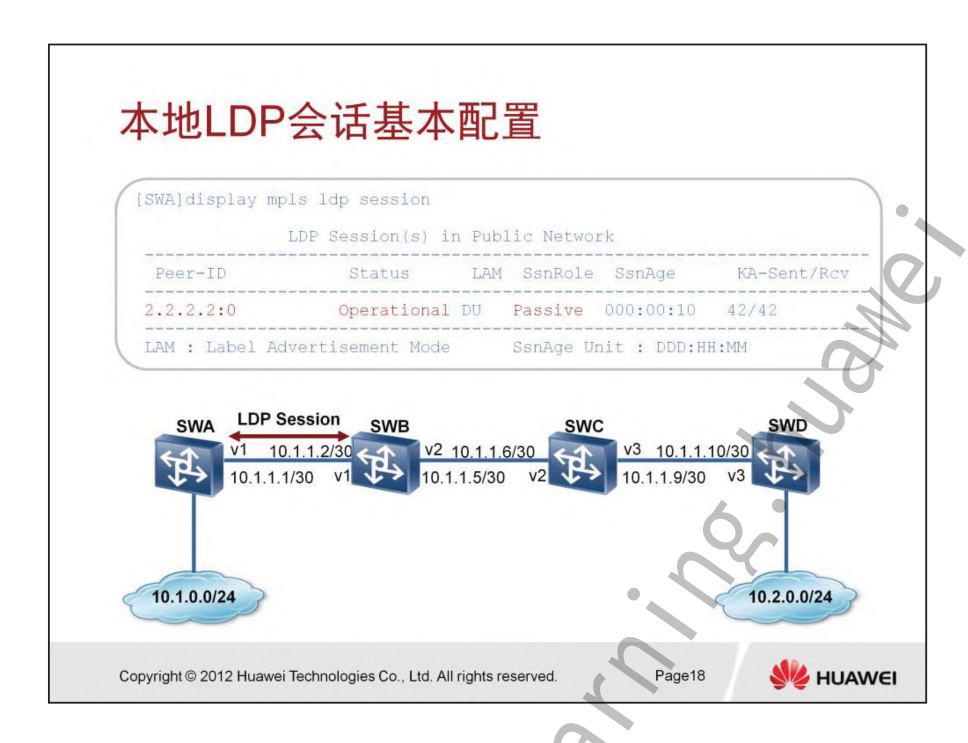
[SWA]inter v1

[SWA-Vlanif1]mpls

[SWA-Vlanif1]mpls ldp

接口模式下使能mpls和ldp。

更多资料获取：http://learning.huawei.com/cn



[SWA]dis mpls ldp session

显示mpls ldp session的状态，状态为Operational表示Session已经正常建立。

Peer-ID为LDP Peer的LDP Identifier，由LDP Peer的lsr-id（这里为2.2.2.2）和表示标签空间的2Bytes（这里为0，表示基于平台的标签空间）构成。

由于SWA的lsr-id地址小于SWB的lsr-id地址，所以SWA为被动方，SSnRole为Passive。

另外可以使用以下命令来查看LDP Peer的信息：

[SWA-Vlanif1]dis mpls ldp peer

```
       LDP Peer Information in Public network
 ------------------------------------------------------------------------
 Peer-ID              Transport-Address  Discovery-Source
 ------------------------------------------------------------------------
 2.2.2.2:0            10.1.1.2           Vlanif1
 ------------------------------------------------------------------------
```

更多资料获取：http://learning.huawei.com/cn

本地LDP会话基本配置

```
[SWA]display mpls ldp session verbose

                 LDP Session(s) in Public Network
 ------------------------------------------------------------------
 Peer LDP ID      : 2.2.2.2:0            Local LDP ID   : 1.1.1.1:0
 TCP Connection   : 1.1.1.1 <- 2.2.2.2
 Session State    : Operational          Session Role   : Passive
 Session FT Flag  : Off                  MD5 Flag       : Off
 Reconnect Timer  : ---                  Recovery Timer : ---

 Negotiated Keepalive Timer          : 45 Sec
 Keepalive Message Sent/Rcvd         : 288/288 (Message Count)
 Label Advertisement Mode            : Downstream Unsolicited
 Label Resource Status(Peer/Local)   : Available/Available
 Session Age                         : 000:01:11  (DDD:HH:MM)

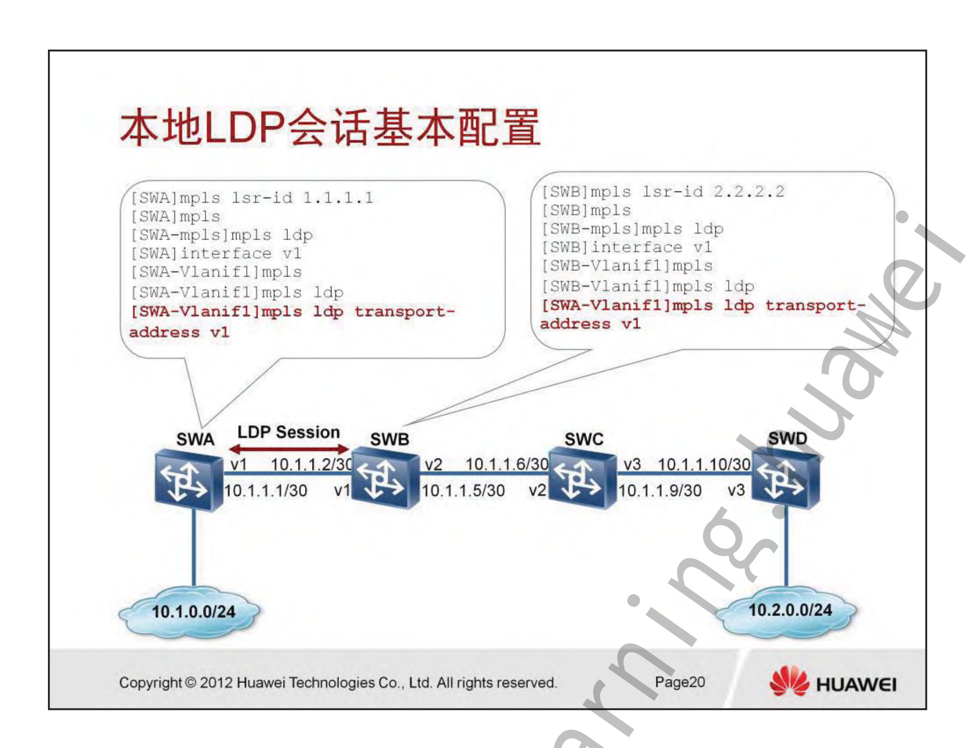
 Addresses received from peer: (Count: 3)
 10.1.1.2            2.2.2.2            10.1.1.5
 ------------------------------------------------------------------
```

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page19 HUAWEI

[SWA]dis mpls ldp session verbose

显示LDP Session的详细信息。从中可以看出VRP在缺省情况下使用lsr-id建立TCP连接。

更多资料获取：http://learning.huawei.com/cn



[SWA-Vlanif1]mpls ldp transport-address v1

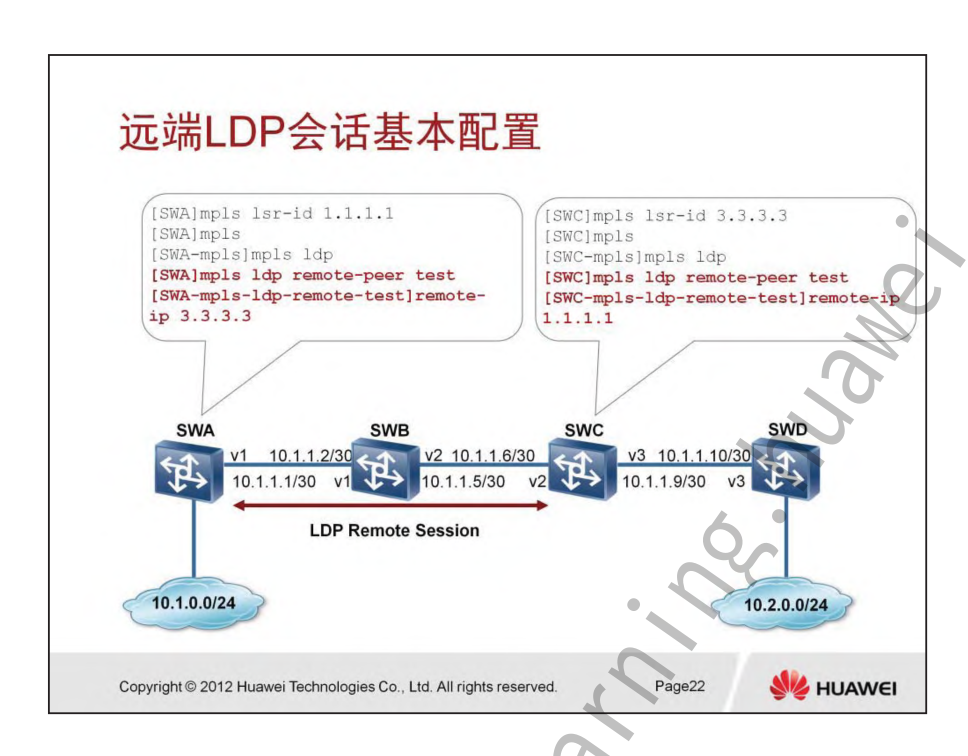
配置使用直连接口v1建立TCP连接。

更多资料获取：http://learning.huawei.com/cn



从LDP Session详细信息看出，在配置Transport Address之后，LDP使用Transport Address配置的IP地址建立TCP连接。

更多资料获取：http://learning.huawei.com/cn



配置SWA和SWC之间建立Remote LDP Session。

配置思路:

1、与本地LDP会话相同，也需要首先配置lsr-id并通过路由协议保证SWA和SWC的lsr-id之间的可达性。

2、配置LDP remote peer。

配置说明:

[SWA]mpls ldp remote-peer test

[SWA-mpls-ldp-remote-test]remote-ip 3.3.3.3

首先创建一个远端对等体，然后指定对等体的lsr-id。

更多资料获取：http://learning.huawei.com/cn



可以看出，SWA和SWC正确建立了LDP Remote Session。

也可以使用以下命令查看LDP Remote Peer的信息：

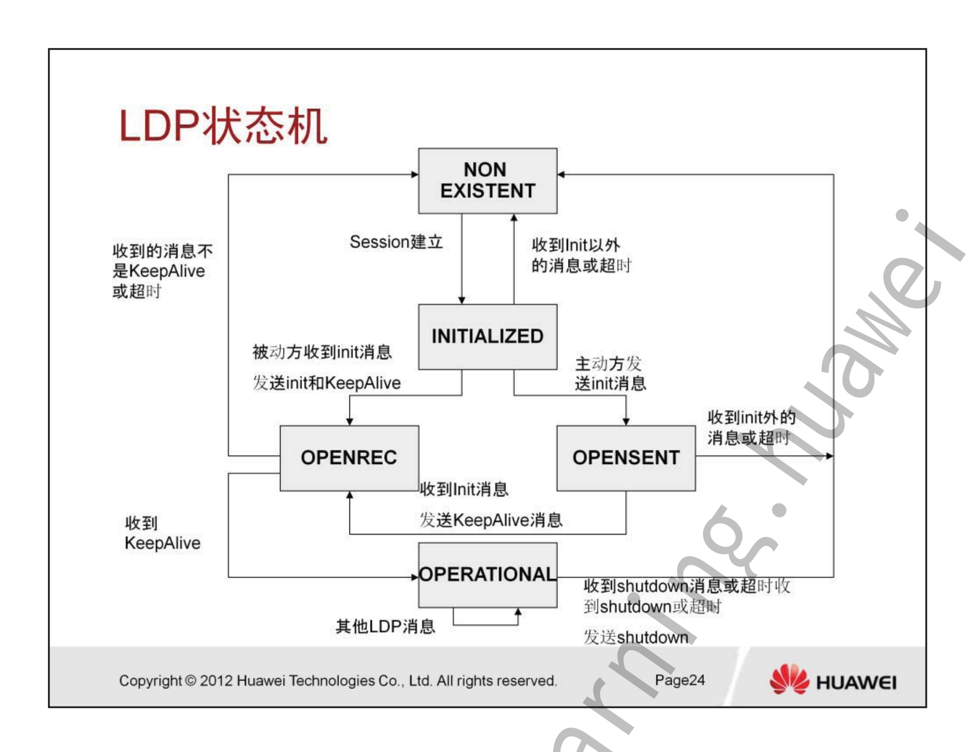
```
[SWA]dis mpls ldp remote-peer test
                 LDP Remote Entity Information
 ------------------------------------------------------------------------
 Remote Peer Name: test
 Remote Peer IP: 3.3.3.3           LDP ID: 1.1.1.1:0
 Transport Address: 1.1.1.1         Entity Status: Active
 Configured Keepalive Timer: 45 Sec    Configured Hello Timer: 45
Sec
 Negotiated Hello Timer: 45 Sec     Hello Packet sent/received: 100/98
```

可以看出，缺省情况下，LDP也是使用lsr-id来建立TCP连接的从而建立LDP Remote Session的（如上所示Transport Address: 1.1.1.1 ）。

可以通过以下命令来修改Transport Address：

```
[SWA]mpls ldp remote-peer test
[SWA-mpls-ldp-remote-test]mpls ldp transport-address ?
  LoopBack  LoopBack interface
```


更多资料获取：http://learning.huawei.com/cn



LDP Session协商过程可以通过状态机来描述。如图所示，有5种状态。分别是NON-EXISTENT，INITIALIZED，OPENREC，OPENSENT，OPERATIONAL。

NON-EXISTENT状态：该状态为LDP Session最初的状态，在此状态双方发送HELLO消息，选举主动方，在收到TCP连接建立成功事件的触发后变为INITIALIZED状态。

INITIALIZED状态：该状态下分为主动方和被动方两种情况，主动方将主动发送Initialization消息，转向OPENSENT 状态，等待回应的Initialization消息；被动方在此状态等待主动方发给自己的Initialization消息，如果收到的Initialization消息的参数可以接受，则发送Initialization和KeepAlive转向OPENREC状态。主动方和被动方在此状态下收到任何非Initialization消息或等待超时时，都会转向NON-EXISTENT状态。

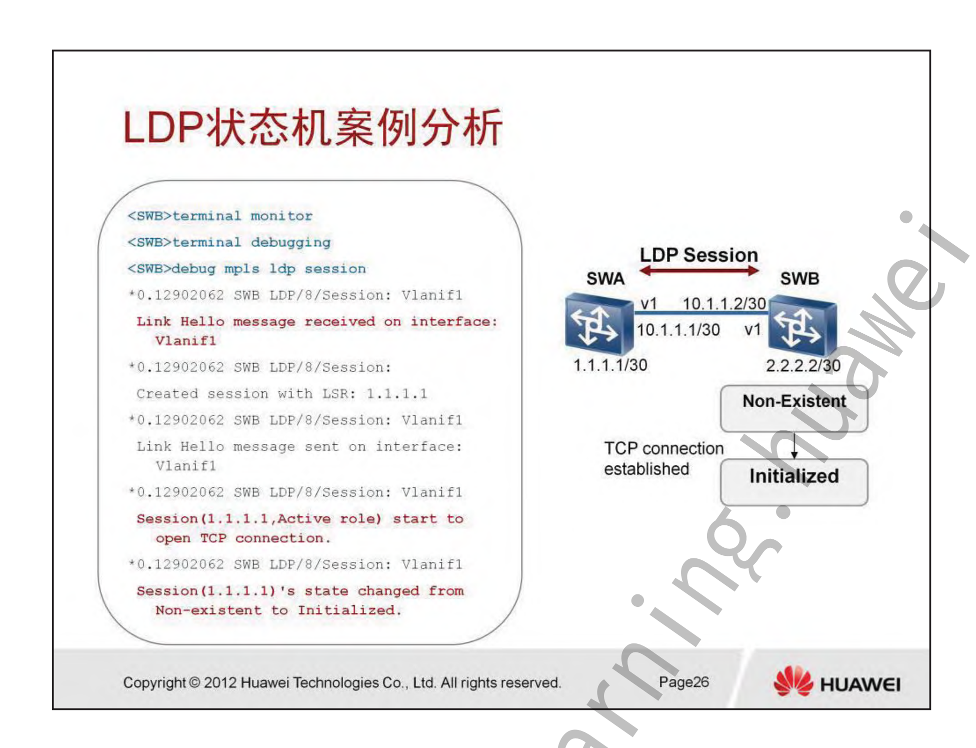
OPENSENT 状态：此状态为主动方发送Initialization消息后的状态，在此状态等待被动方回答Initialization消息和KeepAlive消息，如果收到的Initialization消息中的参数可以接受则转向OPENREC状态，如果参数不能接受或Initialization消息超时则断开TCP连接转向NON-EXISTENT状态。

更多资料获取：http://learning.huawei.com/cn

OPENREC状态：在此状态不管主动方还是被动方都是发出KeepAlive后的状态，在等待对方回应KeepAlive，只要收到KeepAlive消息就转向OPERATIONAL状态；如果收到其它消息或KeepAlive超时则转向NON-EXISTENT状态。

OPERATIONAL状态：该状态是LDP Session成功建立的标志。在此状态下可以发送和接收所有其它的LDP消息。在此状态如果KeepAlive超时或收到致命错误的Notification消息（Shutdown消息）或者自己主动发送Shutdown消息主动结束会话，都会转向NON-EXISTENT状态。

更多资料获取：http://learning.huawei.com/cn



打开mpls ldp session调试信息，进一步了解LDP状态机的迁移过程。

<SWB>terminal monitor

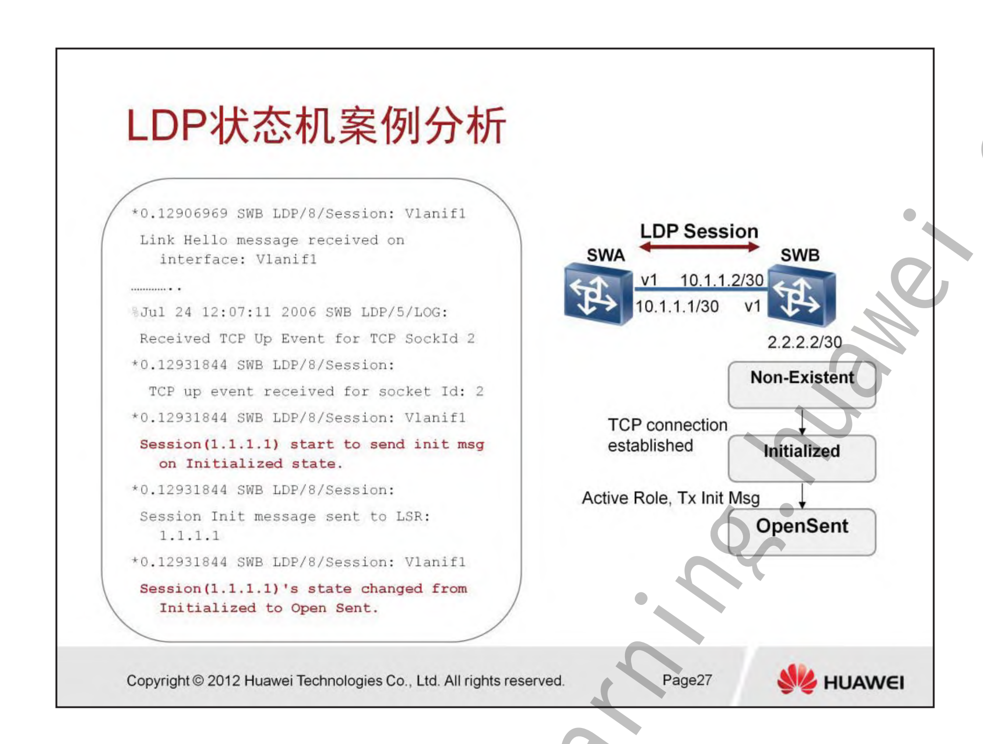
<SWB>terminal debugging

<SWB>debug mpls ldp session

打开mpls ldp session调试开关并将调试信息输出到控制台。

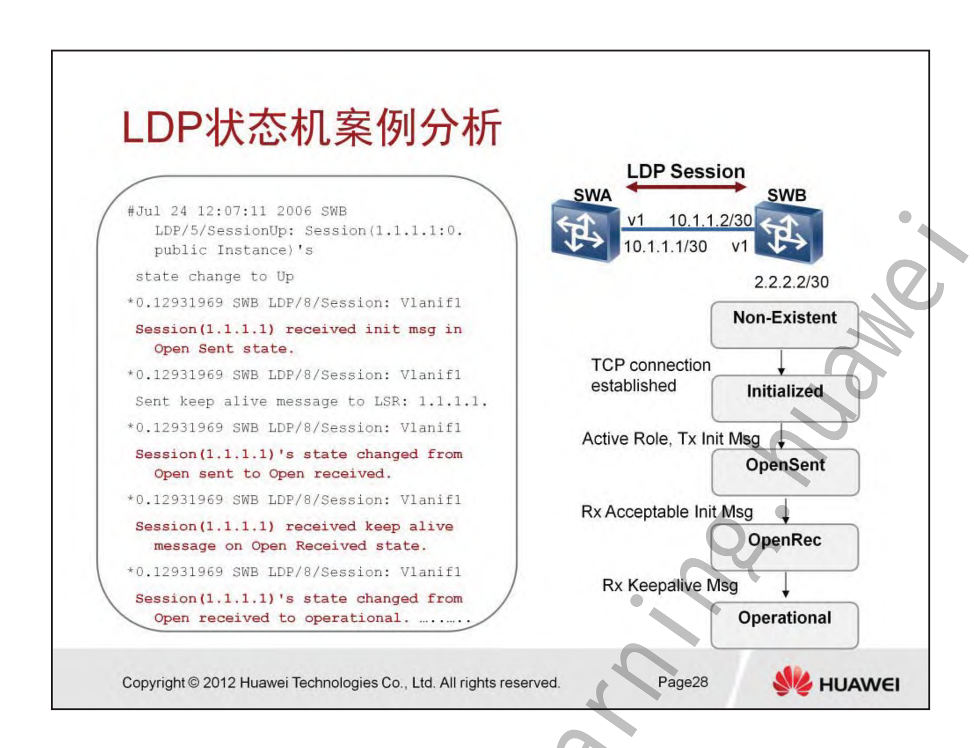
从调试信息可以看出，SWB作为主动方发起TCP连接。状态机从NON-EXISTENT转移到INITIALIZED。

更多资料获取：http://learning.huawei.com/cn



在INITIALIZED状态发送Initialization Message并转移到OPENSent状态。

更多资料获取：http://learning.huawei.com/cn



OPENSent状态接收Initialization Message并发送KeepAlive Message，迁移到OPENRec状态。在OPENRec状态接收到KeepAlive Message后，迁移到OPERATIONAL状态。

更多资料获取：http://learning.huawei.com/cn



更多资料获取：http://learning.huawei.com/cn



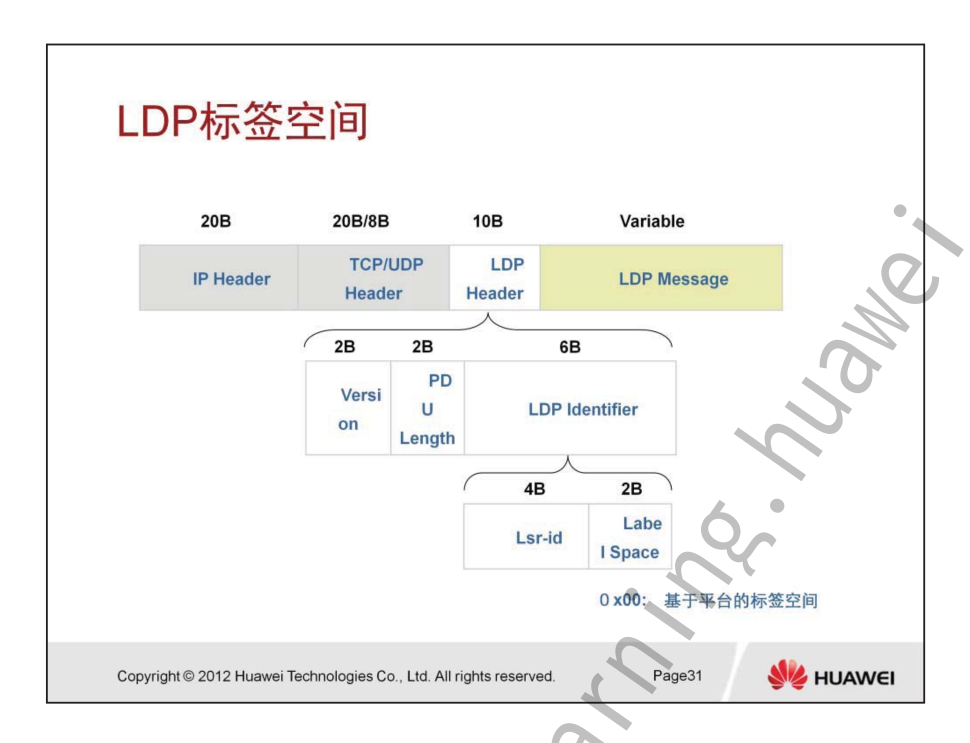
目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page30

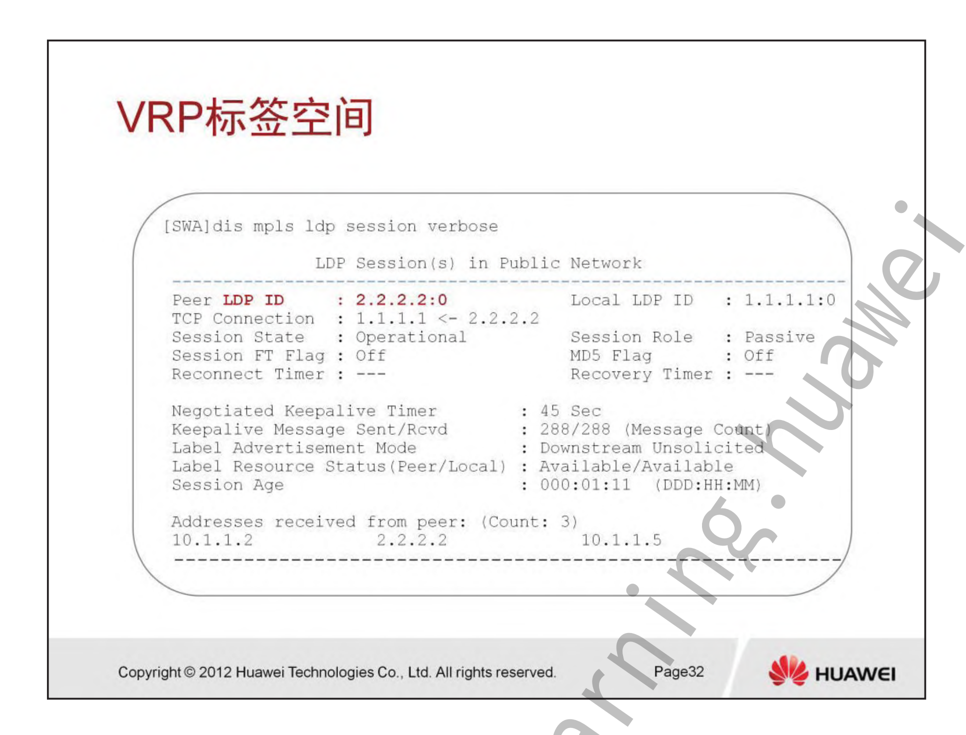


更多资料获取：http://learning.huawei.com/cn



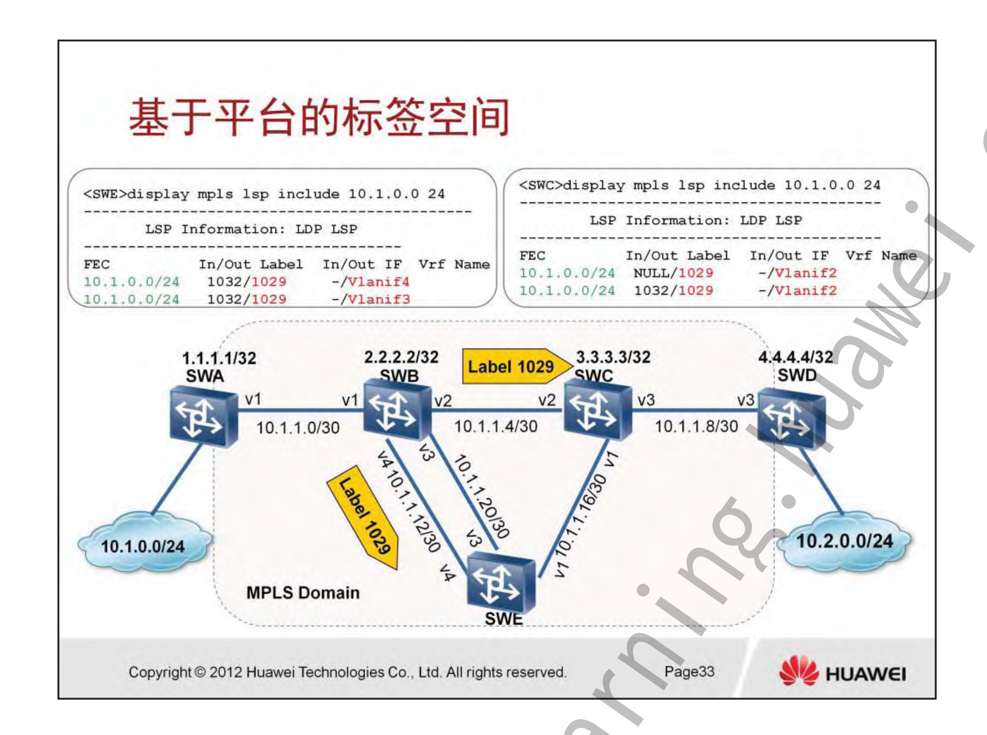
LDP Header中包含6Bytes的LDP Identifier，其中前4Bytes表示Lsr-id，后2Bytes表示标签空间。标签空间有基于平台的标签空间和基于接口的标签空间，当后2Bytes填充0时表示基于平台的标签空间。帧模式封装的MPLS使用基于平台的标签空间，信元模式的MPLS使用基于接口的标签空间。

更多资料获取：http://learning.huawei.com/cn



VRP使用基于平台的标签空间。

更多资料获取：http://learning.huawei.com/cn

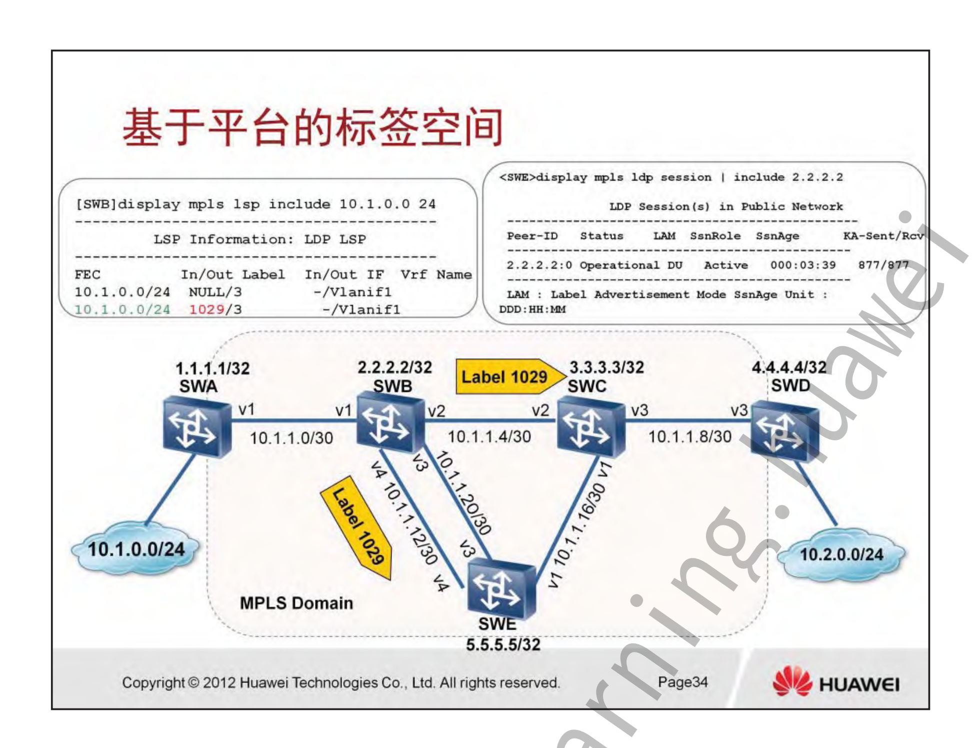


基于平台的标签空间中，LSR为一个目的网段分配只分配一个标签，并将该标签发送给所有的LDP Peers。该标签基于平台，可以用于本LSR任意一个入接口。故该方式可以节省标签。

帧模式的MPLS缺省都使用基于平台的标签空间。

如上图，SWB为10.1.0.0分配一个标签1029并分别发送给他的LDP Peer SWC和SWE。所以SWC上到达10.1.0.0的数据包封装出标签1029并通过接口Vlanif2转发给下一跳SWB，同样SWE上到达10.1.0.0的数据包也封装出标签1029并通过接口Vlanif4和Vlanif3（在SWE和SWB之间存在两条链路）转发给下一跳SWB 。

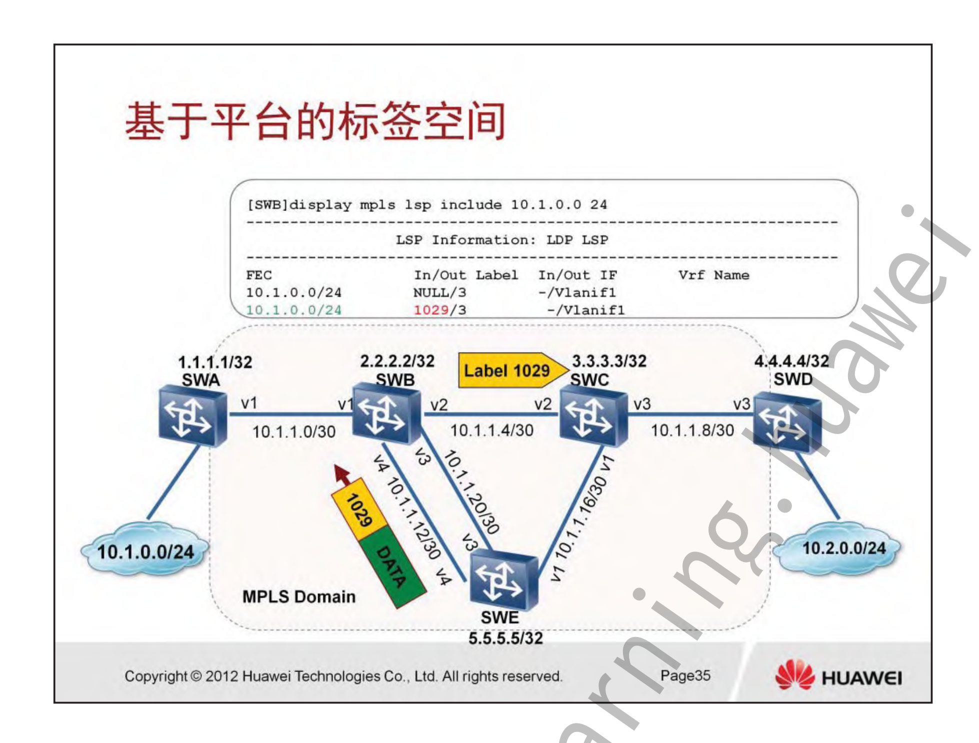
更多资料获取：http://learning.huawei.com/cn



在SWB上根据入标签转发，入标签为1029的数据包从接口Vlanif1转发出去。

另外还可以看出在SWE和SWB之间虽然存在两条链路，但是LDP只建立一个Session传递标签。故基于平台的标签空间可以最小化LDP Session的数量。

更多资料获取：http://learning.huawei.com/cn



但是基于平台的标签空间在安全性上有这样一个弊端：假设SWB只向SWC分发了标签1029，并没有向SWE发送标签，即SWB不希望标签转发从SWE来的数据包。但是如果在SWE上有攻击者伪造了标签为1029的报文发给了SWB，由于SWB转发的时候只是根据入标签去匹配但是并不检查入接口，所以按照SWB上的标签转发表，SWB也会将这个非法报文转发。

更多资料获取：http://learning.huawei.com/cn



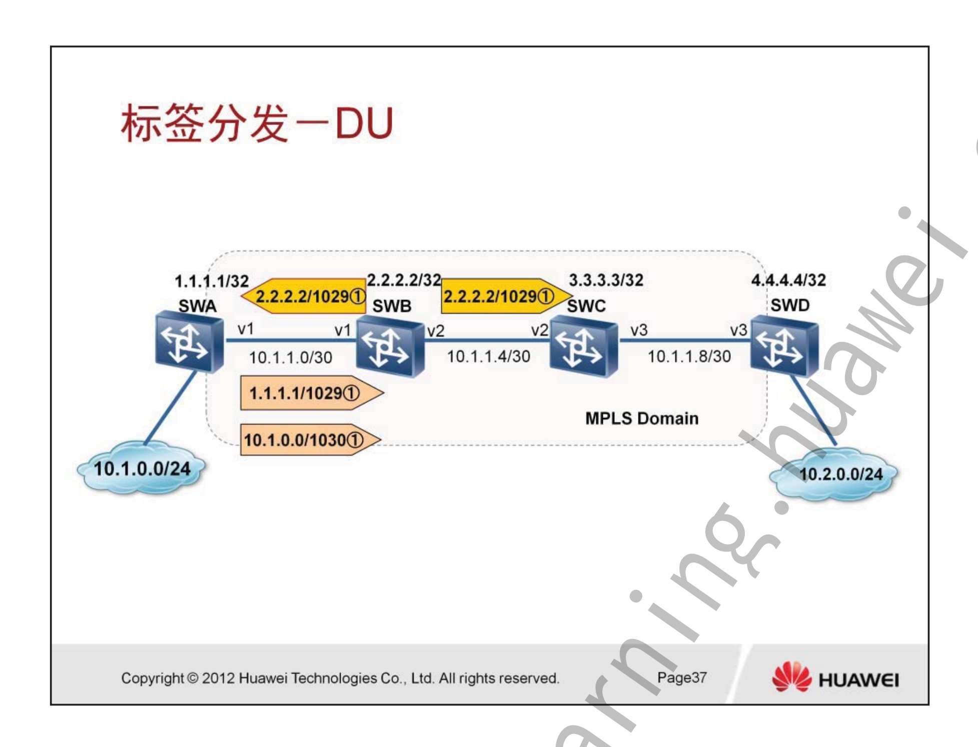
目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

Page36

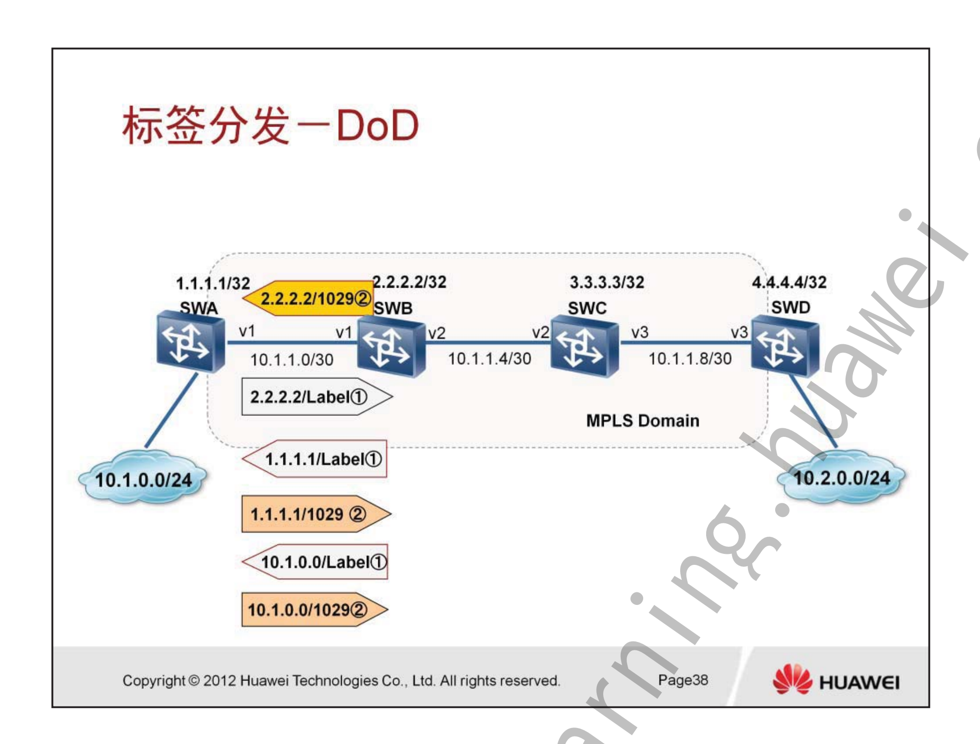
HUAWEI

更多资料获取：http://learning.huawei.com/cn



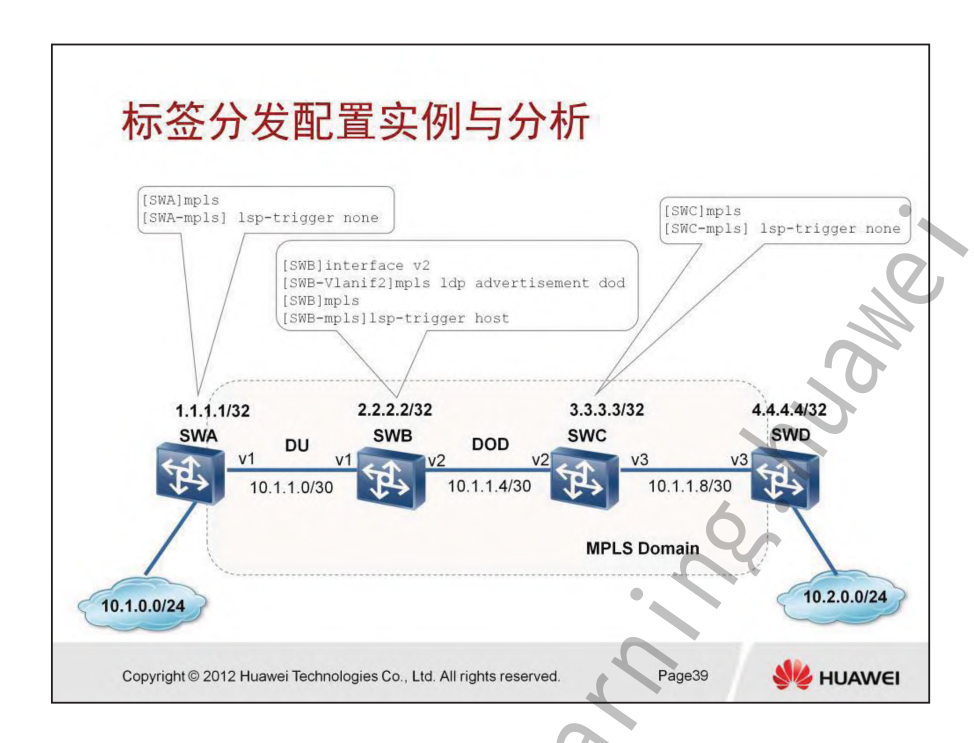
标签分发有两种方式DU（Distribution Unsolicited）和DOD（Distribution on Demand）。DU方式下，无需上游交换机请求，下游LSR将根据某一触发策略向上游LSR发送相应网段（通过配置可以分别实现对路由表中所有网段、对路由表中主机路由或者对特定的IP Prefix分发标签）的标签映射消息（Label Mapping Message）。这里简单介绍上游和下游的概念：对于2.2.2.2/32，下游LSR为SWB，上游LSR为SWA，SWC；对于1.1.1.1/32，下游LSR为SWA，上游LSR为SWB。

更多资料获取：http://learning.huawei.com/cn



DoD方式下，只有当收到上游交换机请求特定网段的标签请求消息（Label Request Message）时，才发送标签映射消息（Label Mapping Message）给上游交换机。如图中SWB只有收到SWA的标签请求消息（2.2.2.2/Label）时才会发送标签映射消息给SWA（2.2.2.2/1029）。

更多资料获取：http://learning.huawei.com/cn



上图中，SWA和SWB之间使用VRP5.30缺省的标签分发方式DU，SWB和SWC之间配置为DoD方式。

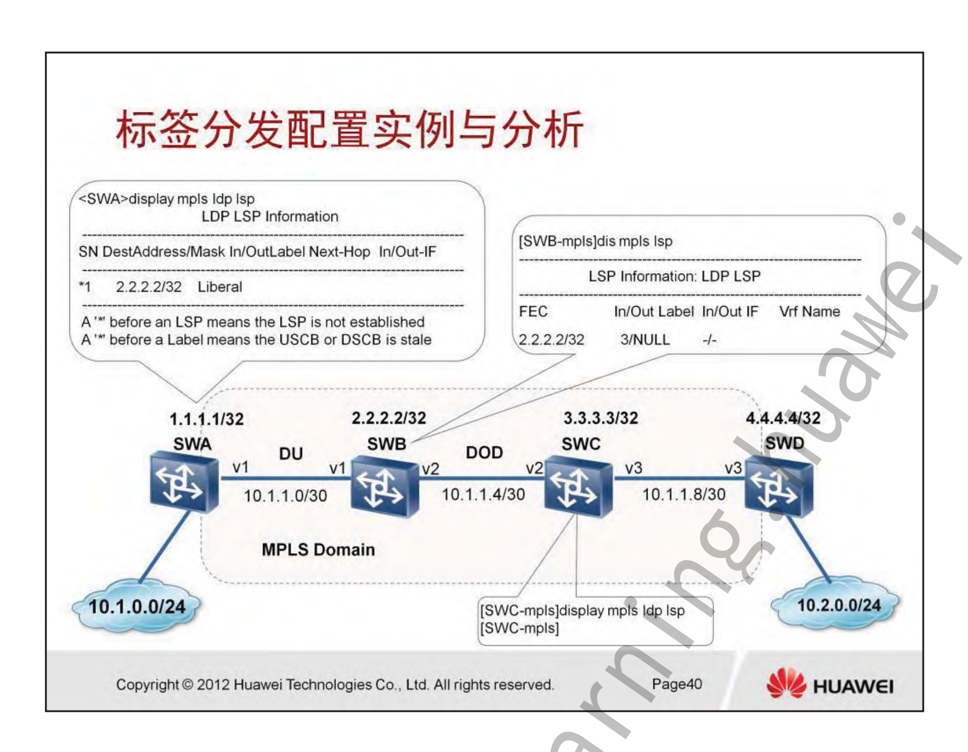
在SWA和SWC上都配置为触发建立LSP，分别观察SWB，SWA和SWC上的标签情况。

配置解释：

[SWB-mpls] lsp-trigger host 配置32位地址的主机IP路由触发建立LSP

[SWC-mpls] lsp-trigger none不触发建立LSP

更多资料获取：http://learning.huawei.com/cn



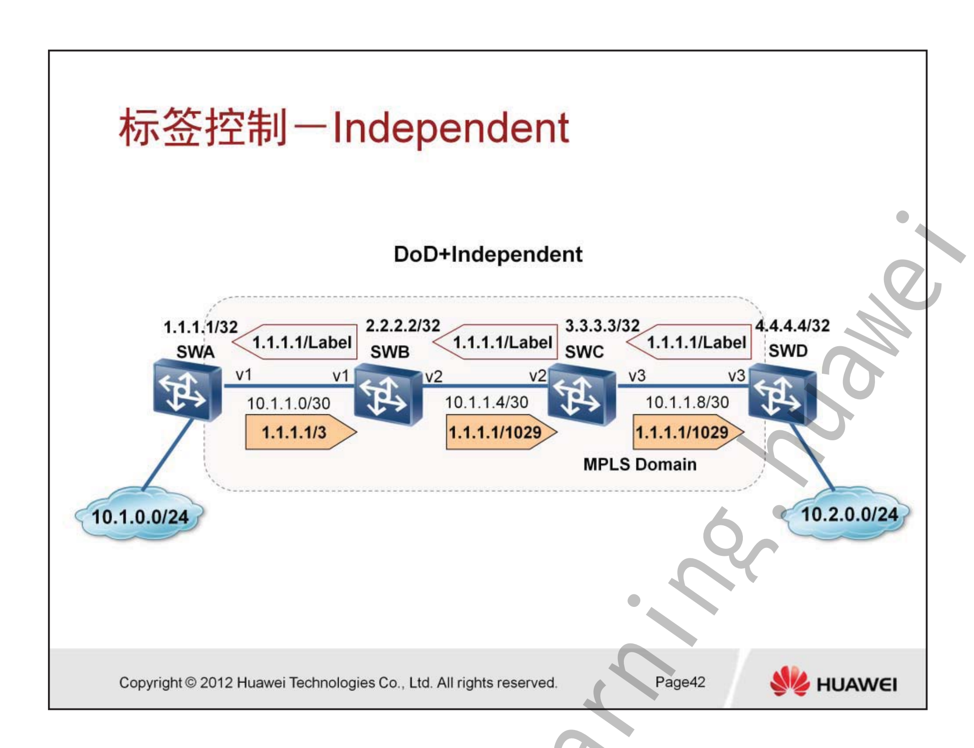
可以看出，由于SWB和SWA之间采用DU方式分发标签，所以虽然上游交换机SWA没有向SWB请求标签，但是SWB依然发送标签映射消息给SWA，在SWA上能够看到SWB分发的标签。

而SWB和SWC之间采用DoD方式分发标签，因为上游交换机SWC没有请求标签，所以SWB没有发送标签映射消息给SWC，故SWC上没有相应的标签映射。

更多资料获取：http://learning.huawei.com/cn



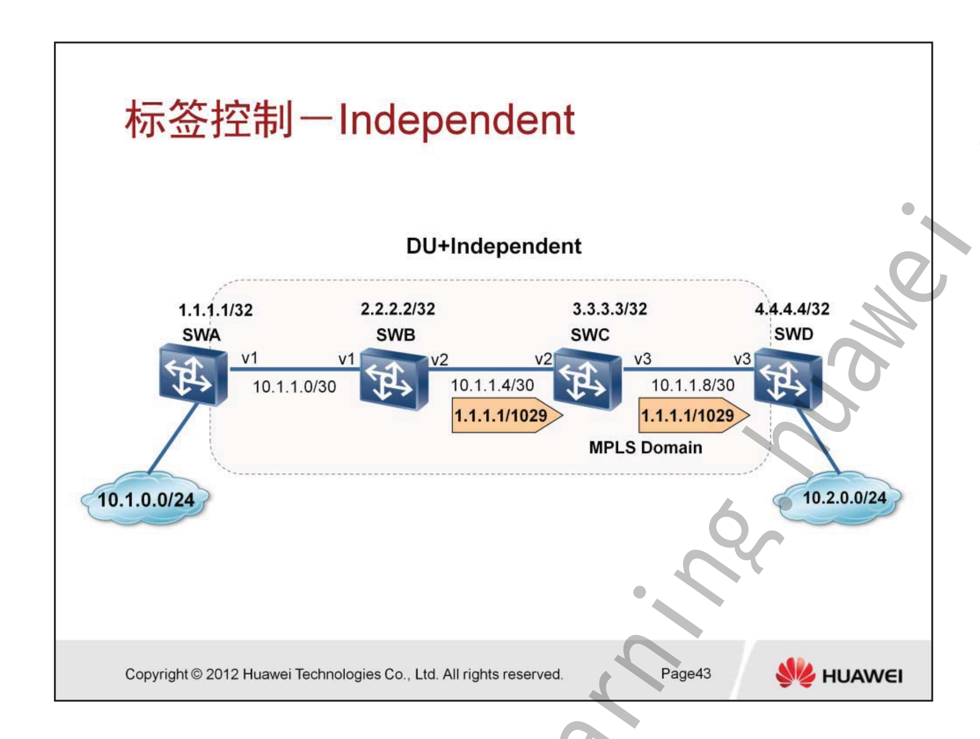
更多资料获取：http://learning.huawei.com/cn



标签控制方式有两种Ordered和Independent。

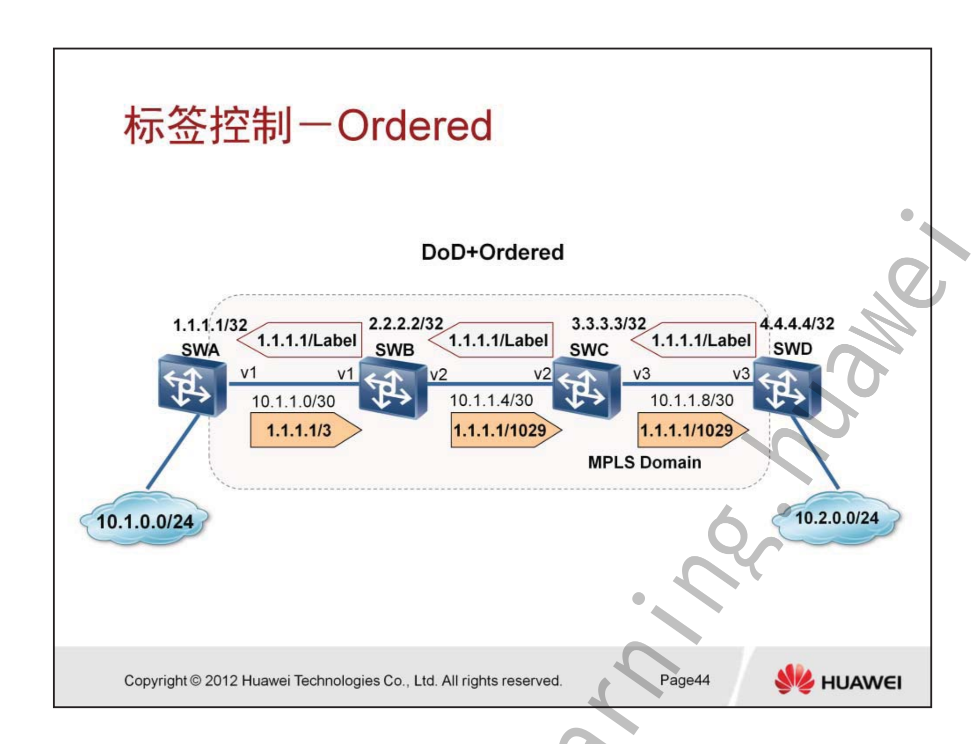
采用Independent控制方式时，每个LSR随时可以向邻居发送标签映射。如，在使用DoD作为标签分发方式的情况下，当LSR（图中SWC）收到上游交换机（图中SWD）发来的标签请求消息时，不必等待自己的下游交换机（如图中SWB）发来标签映射，就可以立即响应该标签请求消息发送自己的标签映射给上游交换机（图中SWD）。

更多资料获取：http://learning.huawei.com/cn



在使用DU作为标签分发方式的情况下，无论何时，只要LSR准备好标签转发相应的FEC，就可以向其邻居发送标签映射。如上图中，配置SWA不触发LSP，SWB对所有路由触发LSP，由于SWB使用Independent作为标签控制方式，所以尽管它的下游交换机（SWA）不会给它分发标签，它也会发送标签映射给它的上游交换机（SWD）。

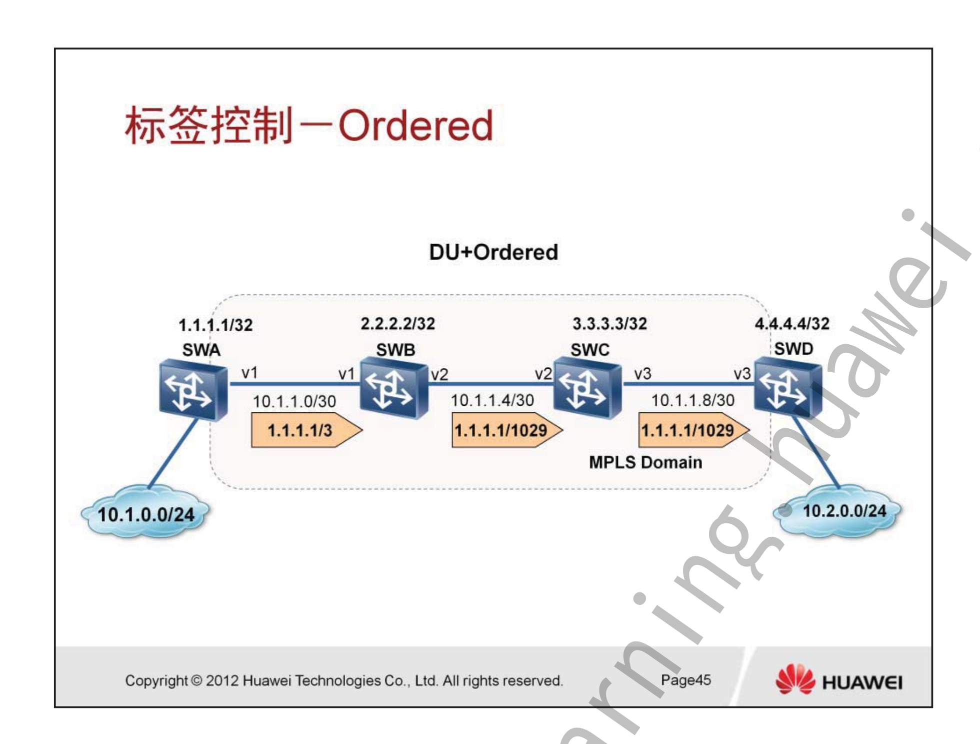
更多资料获取：http://learning.huawei.com/cn



当标签控制方式为Ordered，只有当LSR收到特定FEC下一跳发送的特定FEC-标签映射消息或者LSR是LSP的出口节点时，LSR才可以向上游发送标签映射消息。

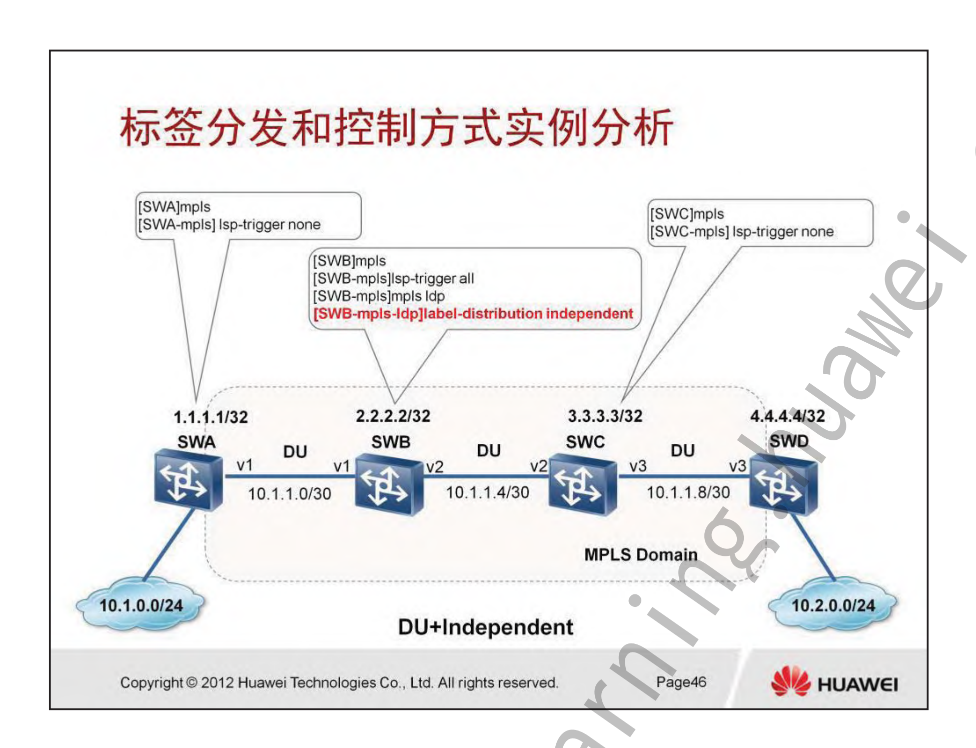
上图给出了标签控制方式为Ordered，分发方式采用DoD时标签的分配情况，SWD向下游交换机（即特定FEC下一跳交换机）LSR SWC请求1.1.1.1的标签，SWC只有在收到他的下游交换机SWB发给他的标签映射消息之后才可能给SWD分发标签，所以SWC在给SWD分发标签之前，首先发送标签请求消息给SWB，SWB再发送标签请求消息给SWA，由于SWA是该LSP的出口节点，所以SWA分发标签给SWB，SWB收到SWA分发的标签后，再分发标签给SWC，SWC收到SWB分配的标签后，再为SWD分发标签。

更多资料获取：http://learning.huawei.com/cn



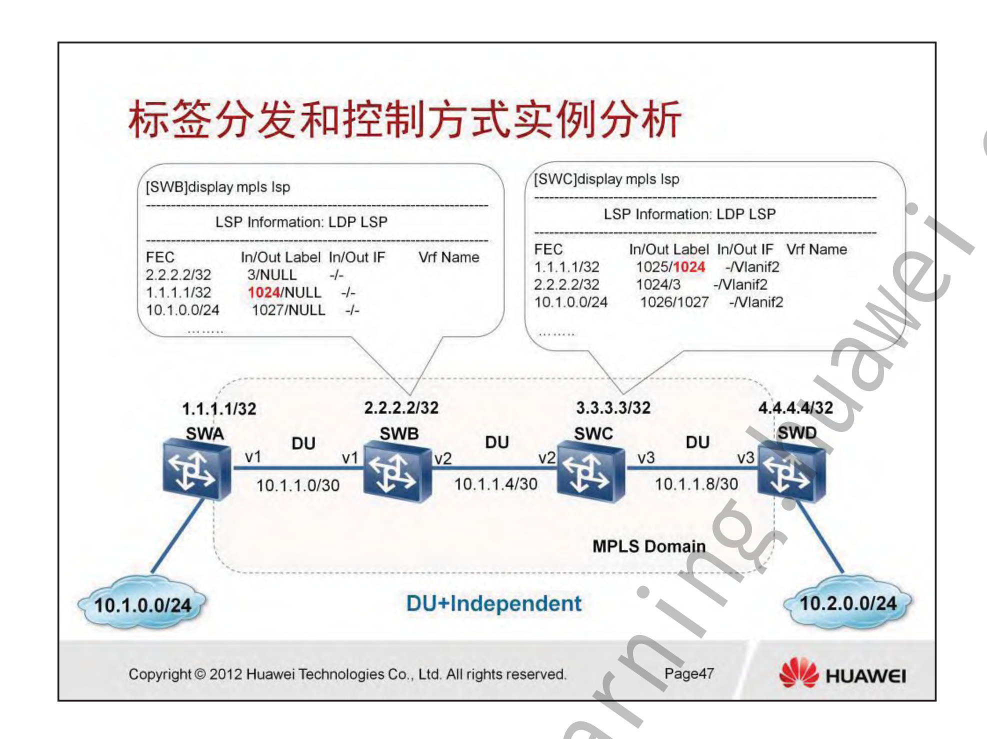
上图给出了标签控制方式为Ordered，标签分发方式为DU时标签的分配情况。下游交换机SWA发送1.1.1.1的标签映射消息给SWB，SWB收到下游交换机SWA分发的标签后给SWC分发标签，SWC收到SWB发来的标签后再给SWD发送标签。

更多资料获取：http://learning.huawei.com/cn



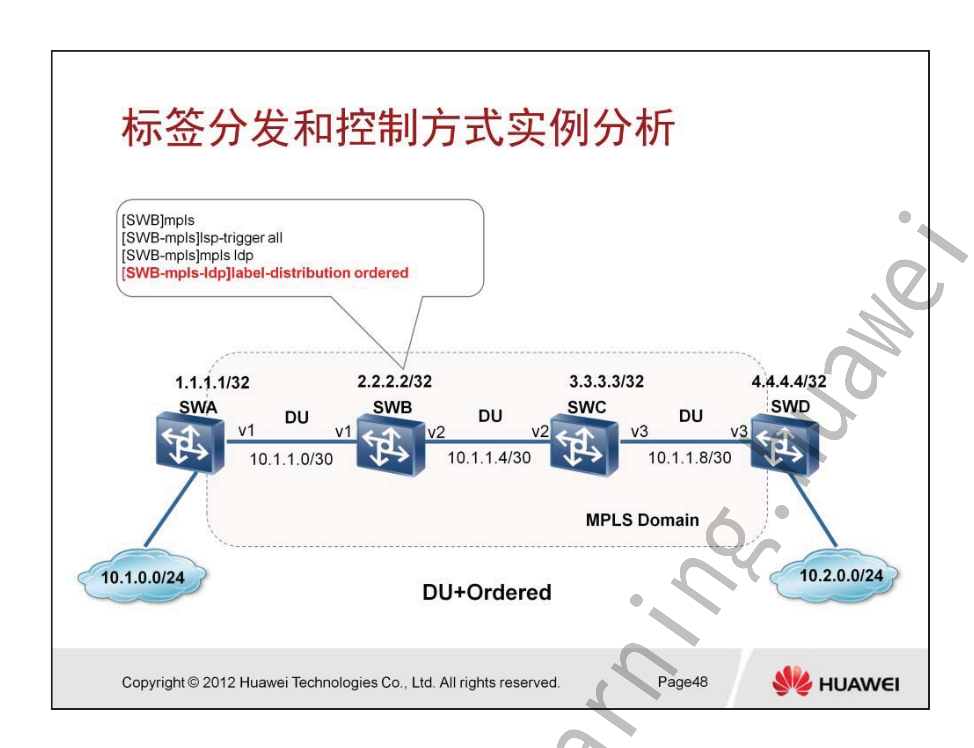
上图中，SWA和SWB之间采用DU标签分发方式，SWB上配置采用Independent标签控制方式，SWB上通过路由协议学习到SWA所连接网段的路由，并且配置SWB所有静态路由和IGP路由项触发建立LSP，SWA配置为不触发LSP 。观察SWB和SWC上的标签信息。

更多资料获取：http://learning.huawei.com/cn



可以看出，虽然SWA（下游交换机）没有分配标签给SWB，但SWB采用Independent控制方式，仍然发送标签映射消息给SWC（上游交换机）。

更多资料获取：http://learning.huawei.com/cn



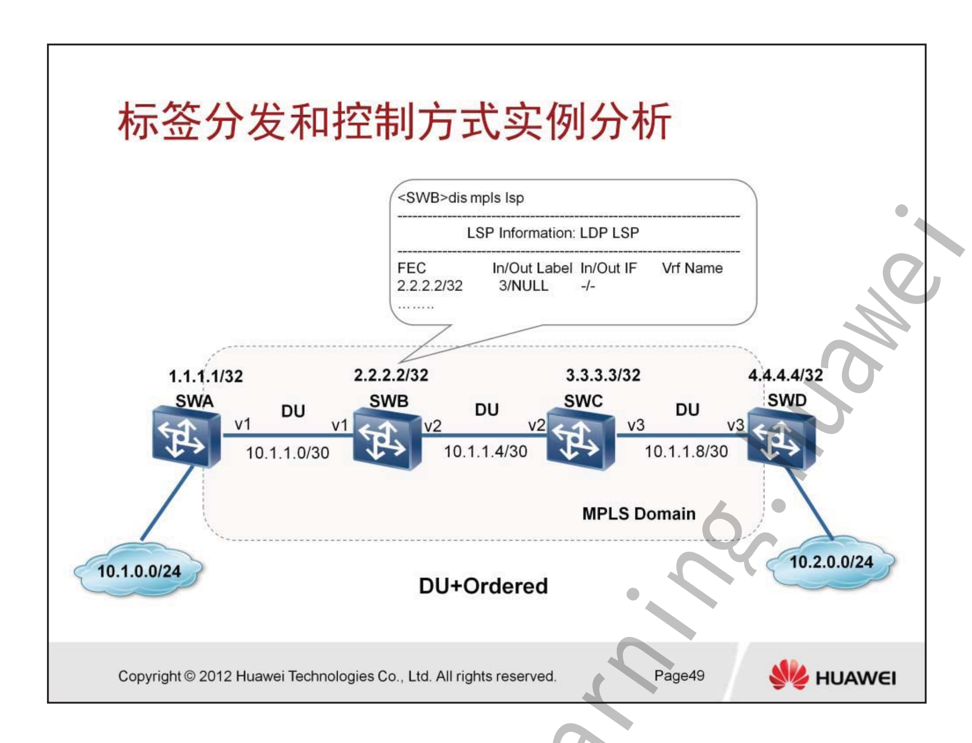
将SWB上的标签控制方式改为Ordered。SWB是否会给1.1.1.1/32分发标签呢?

配置解释:

[SWB-mpls-ldp]label-distribution ordered

配置标签控制方式为Ordered。

更多资料获取：http://learning.huawei.com/cn



使用display mpls lsp查看SWB的LSP，可以看到SWB上配置了使用Ordered的标签控制方式后，由于SWB没有收到下游交换机SWA给1.1.1.1/32分发的标签（在SWA上使用lsp-trigger none命令禁止了触发对本地路由分发标签），所以SWB就不会给上游交换机SWC发送1.1.1.1/32标签映射消息，即不建立LSP。

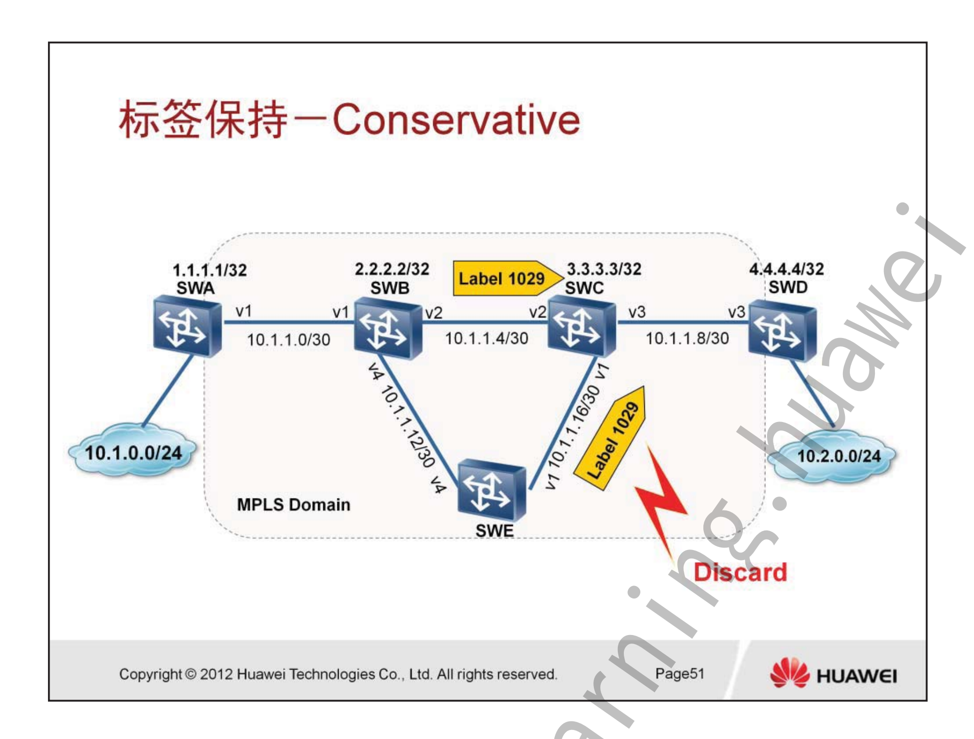
更多资料获取：http://learning.huawei.com/cn



目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

更多资料获取：http://learning.huawei.com/cn

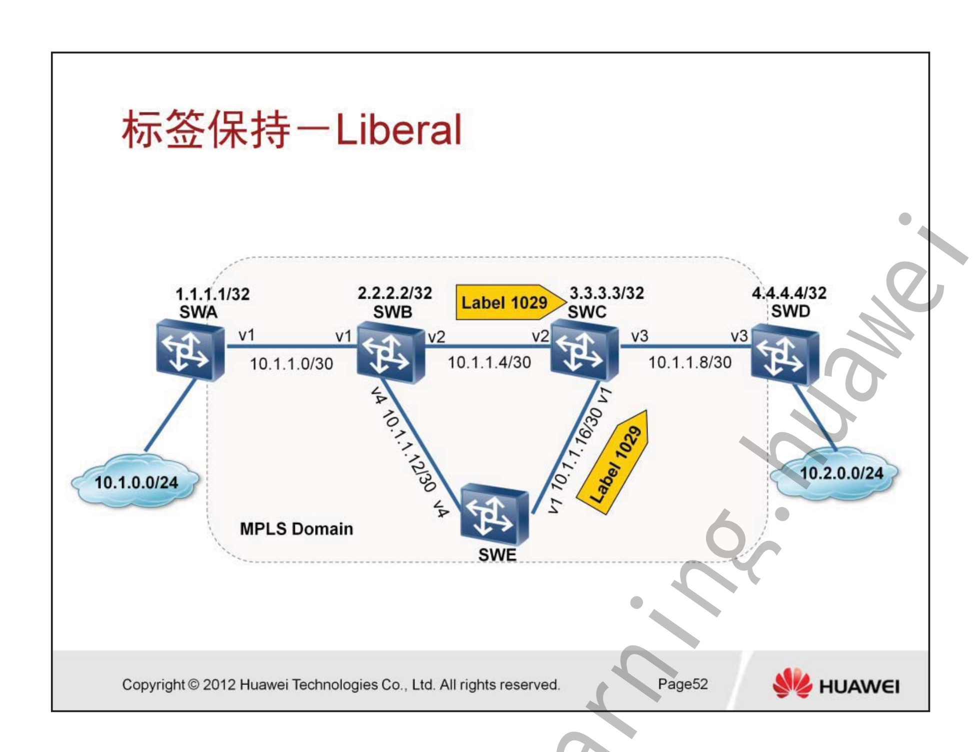


当使用DU标签分发方式时，LSR可能从多个LDP Peer收到到同一网段的标签映射消息，如图中SWC会分别从SWB和SWE收到网段10.1.0.0/24的标签映射消息。如果采用Conservative保持方式，则SWC只保留下一跳SWB发来的标签，丢弃非下一跳SWE发来的标签。

当使用DoD标签分发方式时，如果采用Conservative保持方式， LSR根据路由信息只向它的下一跳请求标签。

Conservative方式的优点在于只需保留和维护用于转发数据的标签，当标签空间有限时，这种方式非常实用，如ATM交换机；缺点在于如果路由表中到达目的网段的下一跳发生了变化，必须从新的下一跳那里获得标签然后才能够转发数据。

更多资料获取：http://learning.huawei.com/cn

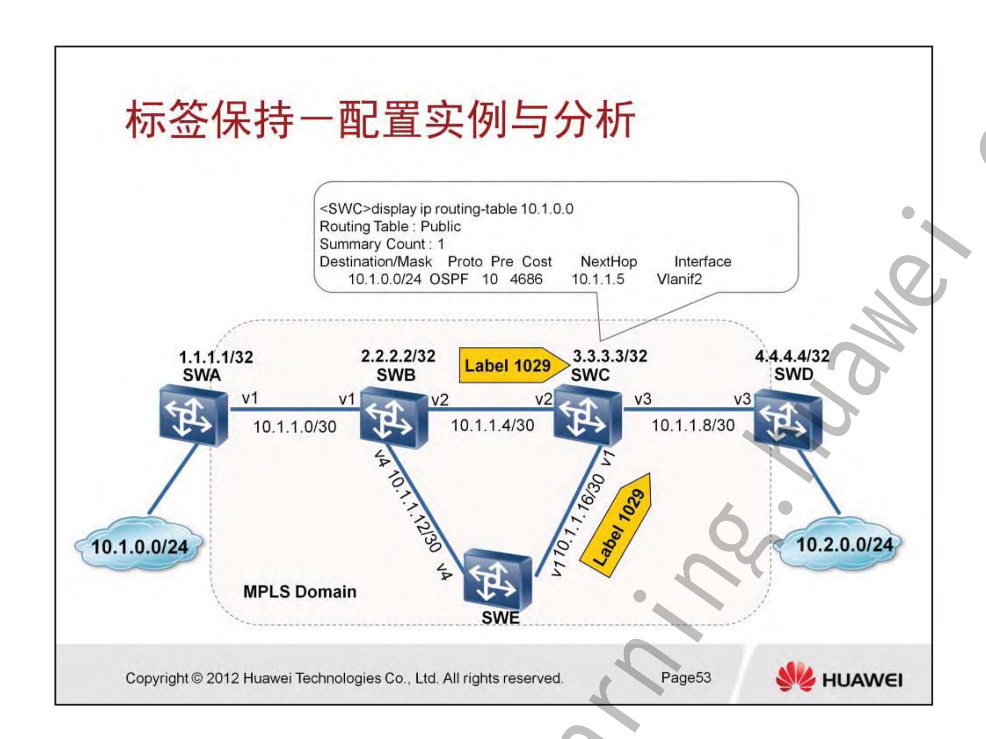


DU标签分发方式下，如果采用Liberal保持方式，则SWC保留所有LDP Peer SWB和SWE发来的标签，无论该LDP Peer是否为到达目的网段的下一跳。

DoD标签分发方式下，如果采用Liberal保持方式， LSR会向所有LDP Peer请求标签。但通常来说，DoD分发方式都会和Conservative保持方式搭配使用。

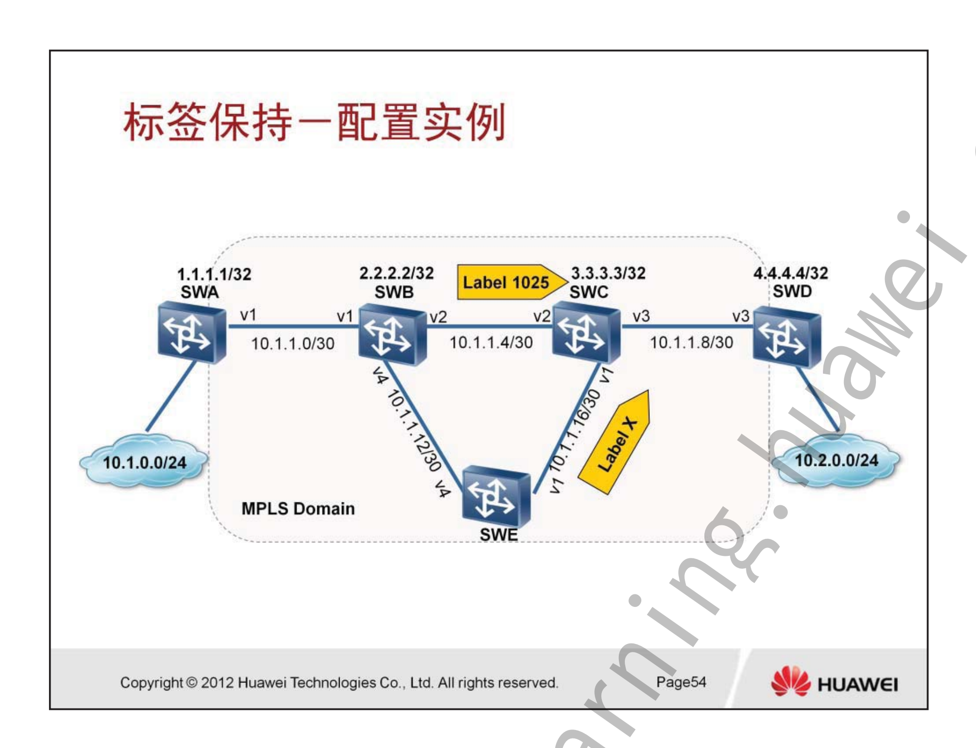
Liberal方式的最大优点在于路由发生变化时能够快速建立新的LSP进行数据转发，因为Liberal方式保留了所有的标签。缺点是需要分发和维护不必要的标签映射。

更多资料获取：http://learning.huawei.com/cn



从中可以看出SWC到10.1.0.0/24的下一跳为SWB。另外，缺省的标签保持方式为Liberal。

更多资料获取：http://learning.huawei.com/cn



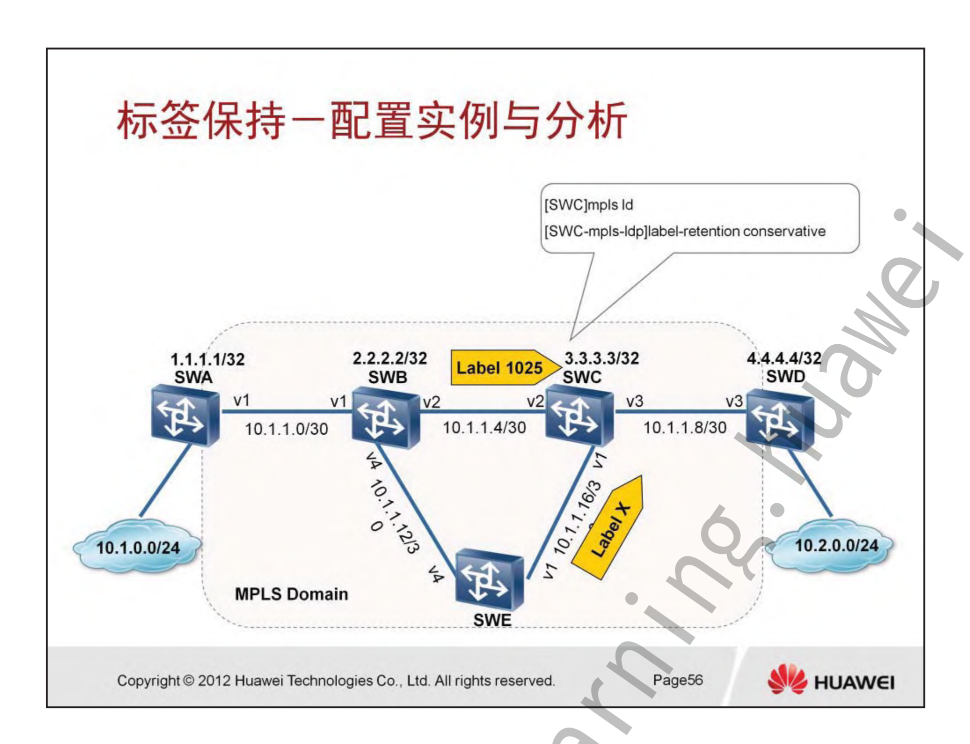
从中可以看出SWC保留了SWB和SWE分发的标签。

更多资料获取：http://learning.huawei.com/cn



从中可以看出SWC保留了SWB和SWE分发的标签。

更多资料获取：http://learning.huawei.com/cn



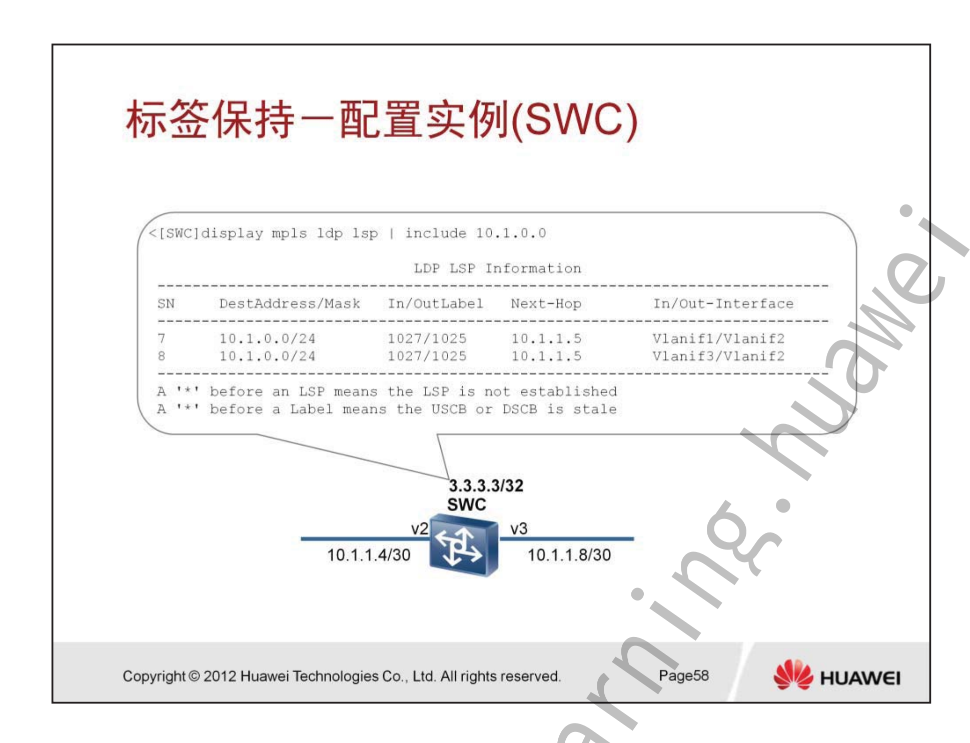
配置SWC采用conservative标签保持方式。

更多资料获取：http://learning.huawei.com/cn



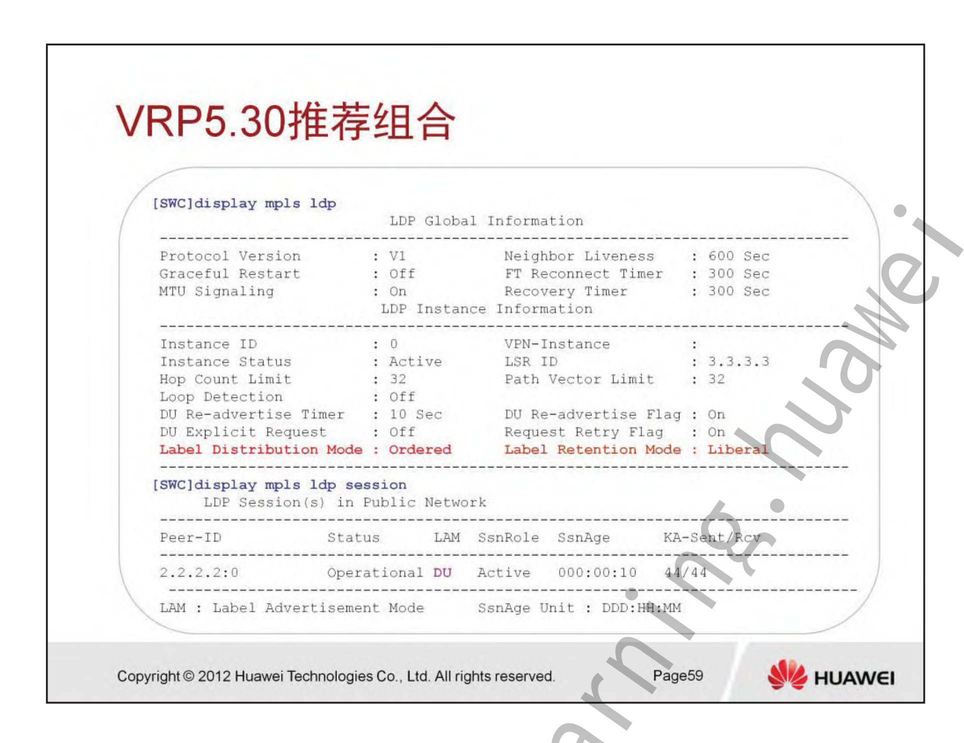
从中可以看出SWC只保留了SWB分发的标签。

更多资料获取：http://learning.huawei.com/cn



从中可以看出SWC只保留了SWB分发的标签。

更多资料获取：http://learning.huawei.com/cn



VRP推荐组合为DU+Ordered+Liberal。该组合为VRP缺省设置。

更多资料获取：http://learning.huawei.com/cn

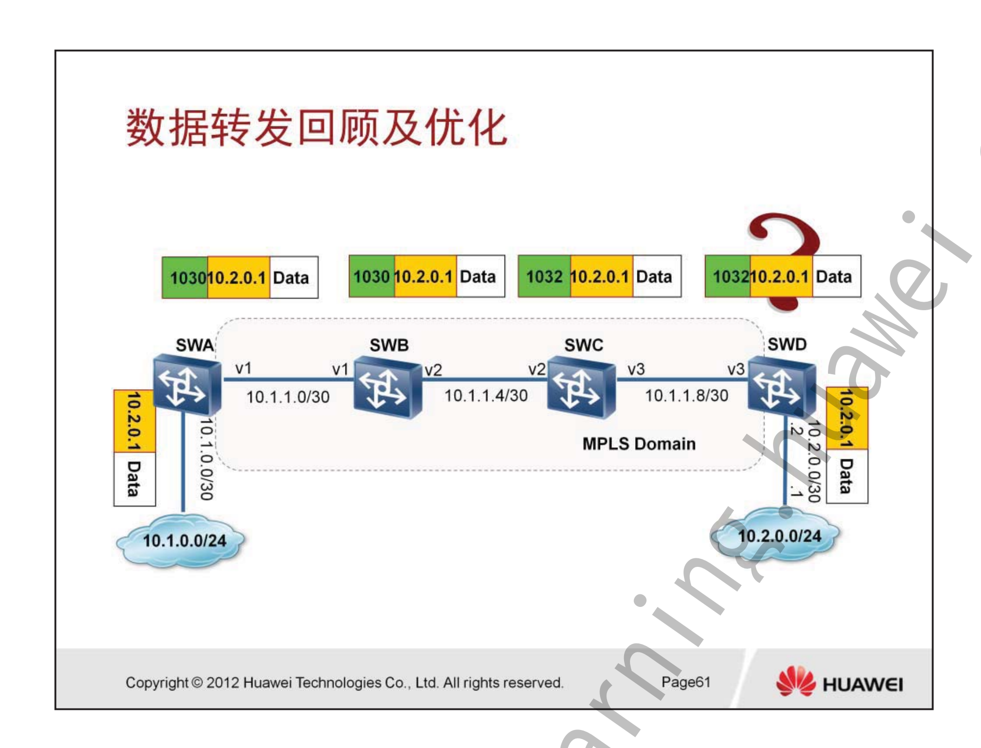


目 录

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved.

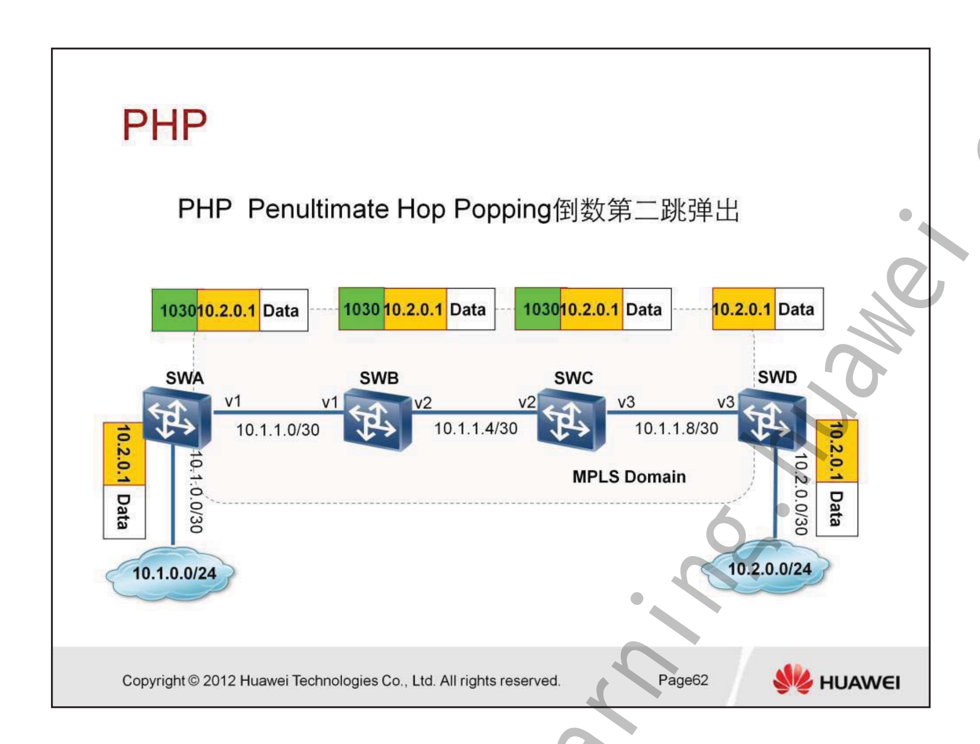
HUAWEI

更多资料获取：http://learning.huawei.com/cn



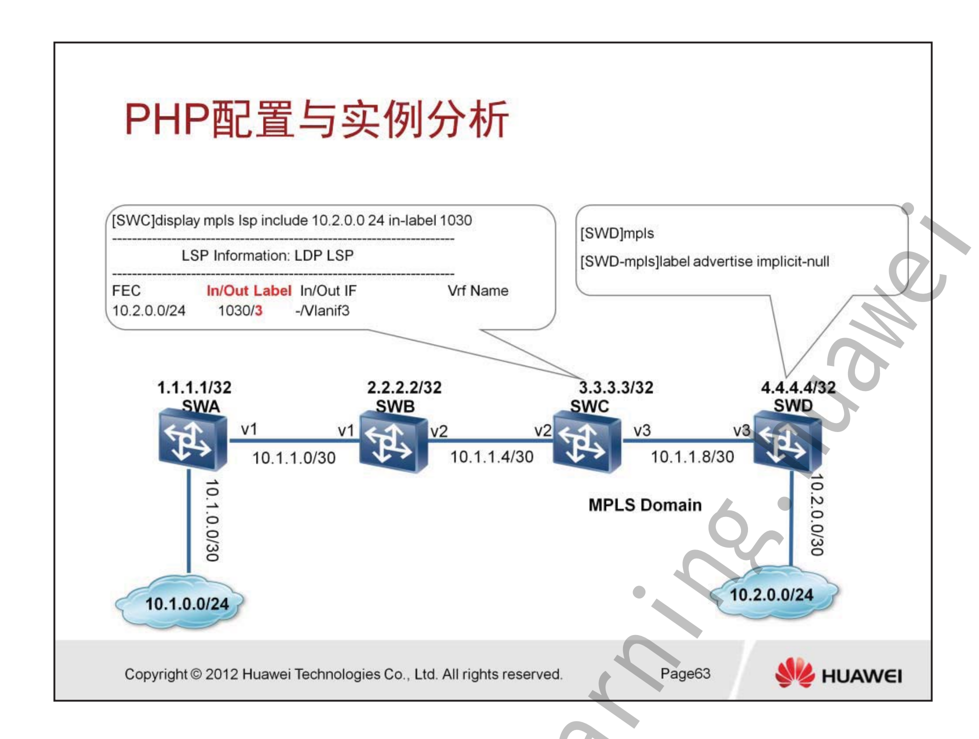
使用LDP完成了标签交互，正确建立LSP后，数据报文就可以沿着LSP转发。这里回顾MPLS标签转发流程。在这个转发过程中Egress LER SWD收到带有标签1032的报文，首先弹出标签，然后根据目的IP地址查找路由表，进行传统的IP转发。实际上对于Egress LER，收到的标签1032对其转发来讲已经没有意义。如果出口LER只进行IP报文的转发而不再分析标签和弹出标签，那么转发的效率会提高。所以可以在倒数第二个LSR SWC上弹出标签后发送IP报文给Egress LER SWD，这样Egress LER SWD就不必处理标签，而是直接转发IP报文到相应目的地即可。这样减少了最后一跳的负担。

更多资料获取：http://learning.huawei.com/cn



PHP（Penultimate Hop Popping）倒数第二跳弹出，可以使得标签在倒数第二跳LSR上弹出。使用倒数第二跳弹出时，倒数第二个LSR依然根据上游LSR标签决定向哪里转发报文，然后直接去掉标签，进行转发，那么当最后一跳LSR（即Egress LER）收到这个报文时，就是传统的IP报文了，这时直接进行传统的IP转发。那么LSR如何知道自己是倒数第二跳呢？倒数第一跳的交换机将为其分配一个特殊的标签3。

更多资料获取：http://learning.huawei.com/cn



LER可以配置3种不同的标签分发方式，以通知倒数第二跳LSR是否应该弹出标签。

[SWD-mpls]label advertise ?

  explicit-null explicit-null

  implicit-null implicit-null

  non-null   non-null

explicit-null为显式空标签，显式空标签值为0。这个值只有出现在标签栈底时才有效，表示报文的标签在分配该标签的这个LSR（即Egress LER）上必须被弹出，然后对此报文进行IP转发；

标签值3表示隐式空标签（implicit-null），这个值不会出现在标签栈中。当一个LSR （倒数第二跳LSR）发现自己被分配了隐式空标签时，它并不用这个值替代栈顶原来的标签，而是直接执行Pop操作。

non-null表示不使用PHP特性，Egress节点向倒数第二跳正常分配标签。缺省情况下，Egress节点向倒数第二跳节点分配隐式空标签implicit-null。

配置解释：

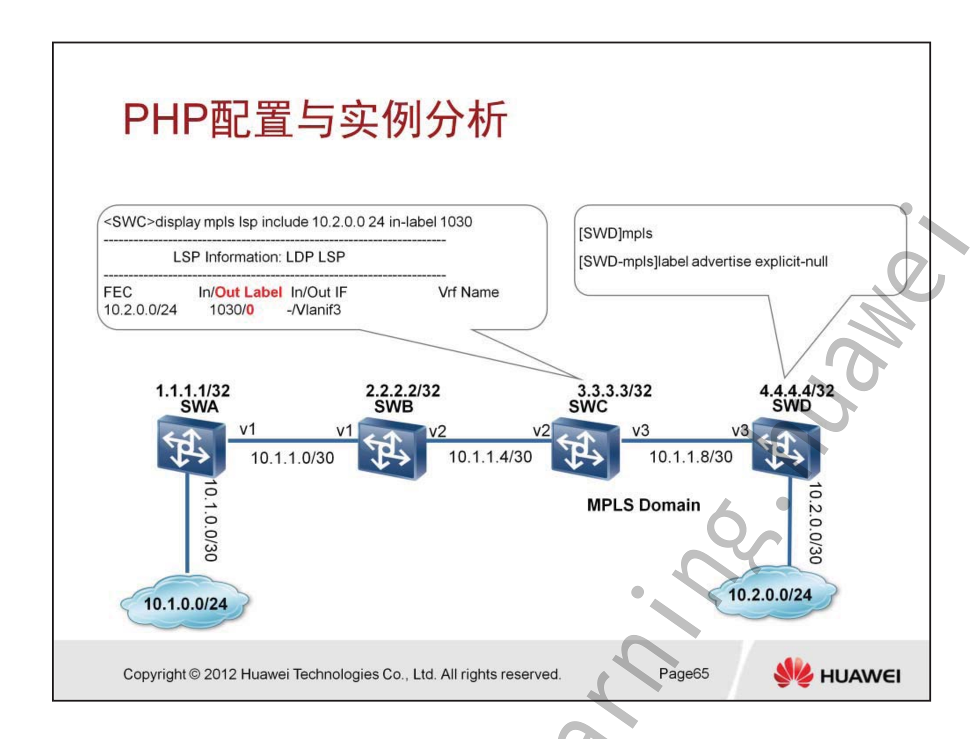
[SWD-mpls]label advertise implicit-null

更多资料获取 http://learning.huawei.com/cn

配置Egress节点向倒数第二跳节点分配隐式空标签implicit-null。该配置为缺省配置。

可以看出SWD为SWC分配了隐式空标签3，SWD收到入标签为1030的报文，将会弹出标签1030，仍然后发送IP报文到SWC。

更多资料获取：http://learning.huawei.com/cn



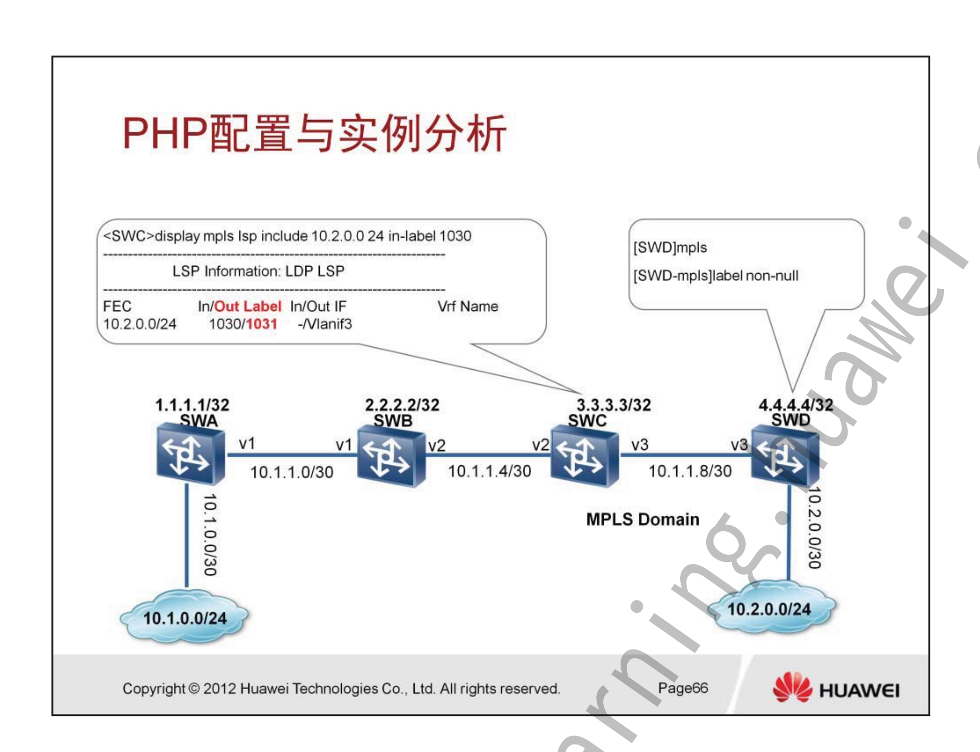
配置解释：

[SWD-mpls]label advertise explicit-null

配置Egress节点向倒数第二跳LSR分配显式空标签explicit-null 。

可以看出，SWD给SWC分配了显式空标签0，SWD收到入标签为0的报文后，由于标签0只在栈底出现，所以弹出该标签，然后进行IP转发。如果是其他普通标签还要判断是否是在栈底，如果不是还要取内层标签通过Mpls转发。

更多资料获取：http://learning.huawei.com/cn



配置解释：

[SWD-mpls]label advertise non-null

表示不使用PHP特性，Egress节点向倒数第二跳正常分配标签。

可以看出，SWD给SWC分配了标签1031。

更多资料获取：http://learning.huawei.com/cn



问 题

LDP邻居发现机制有哪两种，分别有什么区别?

LDP标签分发控制和保持有哪些种方式? 分别有什么区别?

Copyright © 2012 Huawei Technologies Co., Ltd. All rights reserved. Page67



答案:

MPLS转发是根据什么进行数据转发的?

- MPLS是根据标签进行数据转发的。

MPLS常见应用有哪些?

- MPLS VPN，MPLS QoS，MPLS TE。

MPLS封装有哪些方式，各自应用范围是什么?

- 帧模式和信元模式。Ethernet和PPP使用帧模式封装，ATM使用信元模式封装。

LDP邻居发现机制有哪两种，分别有什么区别?

- 基本发现机制和扩展发现机制，基本发现机制用来发现同一链路上的邻居，扩展发现机制用来发现非同一链路上的邻居。

更多资料获取：http://learning.huawei.com/cn

谢谢

www.huawei.com

更多资料获取：http://learning.huawei.com/cn

更多资料获取：http://learning.huawei.com/cn

在线学习资料支持

您可以在华为企业业务网站获得E-Learning课程、培训教材、产品资料、软件工具、技术案例等：

1、 E-Learning课程：登录*华为在线学习网站*，进入“*华为培训/在线学习*”栏目

免费E-Learning课：对网站所有用户免费开放

职业认证E-Learning课：通过任何一项职业认证即可学习所有职业认证培训E-Learning课程

渠道赋能E-Learning课：对华为企业业务合作伙伴免费开放

2、培训教材：登录*华为在线学习网站*，进入“*华为培训/面授培训*”，在具体课程页面即可下载教材。

华为职业认证培训教材、华为产品技术培训教材。无需注册即可下载

3、华为在线公开课(LVC)：http://support.huawei.com/ecommunity/bbs/10154479.html

企业网络、UC&C、安全、存储等诸多领域的职业认证课程，华为讲师公开授课

4、产品资料下载：http://support.huawei.com/enterprise/#tabname=productsupport

5、软件工具下载：http://support.huawei.com/enterprise/#tabname=softwaredownload

更多内容请访问：

- http://learning.huawei.com/cn
- http://support.huawei.com/enterprise/
- http://support.huawei.com/ecommunity/

更多资料获取：http://learning.huawei.com/cn